



INVESTOR IN PEOPLE

GB04/1016

The Patent Office
Concept House
Cardiff Road
Newport
South Wales
NP10 8QQ

REC'D 29 APR 2004

WIPO

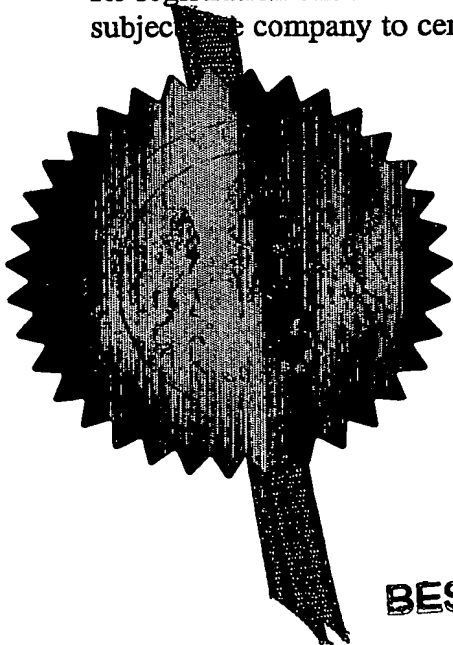
PCT

I, the undersigned, being an officer duly authorised in accordance with Section 74(1) and (4) of the Deregulation & Contracting Out Act 1994, to sign and issue certificates on behalf of the Comptroller-General, hereby certify that annexed hereto is a true copy of the documents as originally filed in connection with the patent application identified therein.

In accordance with the Patents (Companies Re-registration) Rules 1982, if a company named in this certificate and any accompanying documents has re-registered under the Companies Act 1980 with the same name as that with which it was registered immediately before re-registration save for the substitution as, or inclusion as, the last part of the name of the words "public limited company" or their equivalents in Welsh, references to the name of the company in this certificate and any accompanying documents shall be treated as references to the name with which it is so re-registered.

In accordance with the rules, the words "public limited company" may be replaced by p.l.c., plc, P.L.C. or PLC.

Re-registration under the Companies Act does not constitute a new legal entity but merely subjects the company to certain additional company law rules.



Signed

Stephen Handley

Dated 6 April 2004

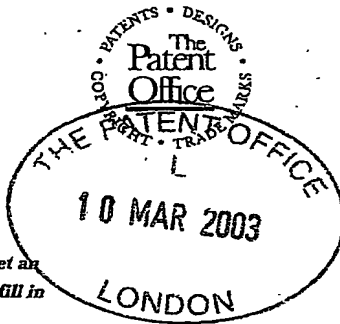
BEST AVAILABLE COPY

**PRIORITY
DOCUMENT**

SUBMITTED OR TRANSMITTED IN
COMPLIANCE WITH RULE 17.1(a) OR (b)

Patents Form 1/77

Patents Act 1977
(Rule 16)



1/77
11MAR83 2791059-2000192
P01/7700 0-00-0305448.3

Request for grant of a patent

(See the notes on the back of this form. You can also get an explanatory leaflet from the Patent Office to help you fill in this form)

The Patent Office

Cardiff Road
Newport
South Wales
NP10 8QQ

1. Your reference P88054 SER

2. Patent application number
(The Patent Office will fill in this part)

110 MAR 2003

0305448.3

3. Full name, address and postcode of the or of each applicant (underline all surnames)

TCP INNOVATIONS LIMITED
9 St Johns Street
Duxford
Cambridge CB2 4RA

Patents ADP number (if you know it)

If the applicant is a corporate body, give the country/state of its incorporation

Cambridge, United Kingdom

858404700

4. Title of the invention IMMUNOASSAY

5. Name of your agent (if you have one)

J.A. KEMP & CO.

"Address for service" in the United Kingdom to which all correspondence should be sent (including the postcode)

14 South Square
Gray's Inn
London
WC1R 5JJ

Patents ADP number (if you know it)

26001

6. If you are declaring priority from one or more earlier patent applications, give the country and the date of filing of the or of each of these earlier applications and (if you know it) the or each application number

Country

Priority application number
(if you know it)

Date of filing
(day / month / year)

7. If this application is divided or otherwise derived from an earlier UK application, give the number and the filing date of the earlier application

Number of earlier application

Date of filing
(day / month / year)

8. Is a statement of inventorship and of right to grant of a patent required in support of this request? (Answer 'Yes' if

Yes

- a) any applicant named in part 3 is not an inventor, or
 - b) there is an inventor who is not named as an applicant, or
 - c) any named applicant is a corporate body.
- See note (d))

Patents Form 1/77

9. Enter the number of sheets for any of the following items you are filing with this form. Do not count copies of the same document

Continuation sheets of this form	-
Description	45
Claim(s)	4
Abstract	- DL
Drawing(s)	44

10. If you are also filing any of the following, state how many against each item.

Priority documents	-
Translations of priority documents	-
Statement of inventorship and right to grant of a patent (Patents Form 7/77)	-
Request for preliminary examination and search (Patents Form 9/77)	-
Request for substantive examination (Patents Form 10/77)	-
Any other documents (please specify)	-

11. I/We request the grant of a patent on the basis of this application.

Signature	J.A. Kemp & Co	Date	10 March 2003
J.A. KEMP & CO.			

12. Name and daytime telephone number of person to contact in the United Kingdom

ROQUES, Sarah Elizabeth
020 7405 3292

Warning

After an application for a patent has been filed, the Comptroller of the Patent Office will consider whether publication or communication of the invention should be prohibited or restricted under Section 22 of the Patents Act 1977. You will be informed if it is necessary to prohibit or restrict your invention in this way. Furthermore, if you live in the United Kingdom, Section 23 of the Patents Act 1977 stops you from applying for a patent abroad without first getting written permission from the Patent Office unless an application has been filed at least 6 weeks beforehand in the United Kingdom for a patent for the same invention and either no direction prohibiting publication or communication has been given, or any such direction has been revoked.

Notes

- If you need help to fill in this form or you have any questions, please contact the Patent Office on 08459 500505.*
- Write your answers in capital letters using black ink or you may type them.*
- If there is not enough space for all the relevant details on any part of this form, please continue on a separate sheet of paper and write "see continuation sheet" in the relevant part(s). Any continuation sheet should be attached to this form.*
- If you have answered 'Yes' Patents Form 7/77 will need to be filed.*
- Once you have filled in the form you must remember to sign and date it.*
- For details of the fee and ways to pay please contact the Patent Office.*

IMMUNOASSAY

The present invention relates to methods of assaying the levels of proteins or antibodies in a test sample. More particularly, methods are provided which allow the relative concentration of many proteins in a pair of samples to be rapidly determined. Further methods are provided which generate a profile of the array of antibodies present in a test sample.

Background to the Invention

10

Increasingly, scientific advances and technological applications are depending on the capability to measure many different parameters about a complex system, such as a living cell, simultaneously. The first examples to become widely available in biology of such "holistic" analyses came from the introduction of "gene chips" which could analyse the levels of gene expression for many hundreds or thousands of genes simultaneously. This technology, which underpins the field of genomics (the study of the co-ordinate regulation of all the genes in the organism), is now ubiquitous and has brought a number of benefits to science and technology.

20

However, genomics is not the only "omics" - the term given to branches of sciences devoted to examining the co-regulation of parameters within a complex system.

Proteomics is the term given to the study of the regulation of all the proteins present in a cell, tissue or biological sample. Metabonomics is the analogous study of all the non-protein (usually low molecular weight) metabolites, such as sugars and fats, in a cell, tissue or biological sample. Both proteomics and metabonomics have been shown to be useful for diagnosing human diseases much more powerfully than the conventional approach of measuring just a few candidate disease markers (such as measuring cholesterol levels to diagnose the presence of heart disease).

30

The utility of "omics" approaches to understanding complex systems (such as human beings) is limited by the ease and robustness of the underpinning technology. For

example, it was the introduction of commercially available gene-chips that led the current rash of genomics research and technology.

In genomics, the gene array tools currently available are relatively easy to use, although they require certain small and relatively cheap specialist pieces of equipment which need to be installed and maintained. Unfortunately, the results obtained are not particularly robust, with coefficient of variations for repeated measures often exceeding 25%. Such inaccuracy severely hampers the use of gene array technology in many, if not all, applications.

Conversely, in metabonomics the tools currently available (such as NMR and IR spectroscopy or mass spectrometry) are inherently robust, often producing repeated-measures coefficients of variation below 2%. However, they are intrinsically complex technologies requiring not only significant capital investment (an NMR machine, for example, may cost in excess of half a million pounds) but also extensive specialist knowledge to operate in a useful way.

Proteomics currently lies somewhere between these two extremes: the technology is somewhat accessible and somewhat robust. Currently, the approaches to proteomics fall into two broad groups: separation based techniques and whole sample techniques.

Considering the separation-based techniques first, the two most commonly used separation technologies are gel electrophoresis and tandem liquid chromatography.

In both cases, the protein mixture is separated into components, which are then analysed by electrospray tandem mass spectrometry to identify the component. These techniques require relatively specialist and capital intensive equipment, and they produce data with repeated measures coefficients of variation down to 10%. Neither technique, however, is well suited to high throughput applications and the amount of data processing required for a single sample is often very large indeed.

The whole sample approach has the advantage of being intrinsically more suited to high throughput applications, such as clinical diagnostics. Unfortunately, the current approaches (of which the best established is the shotgun tandem mass spectrometry approach in which the entire sample is fragmented and then the sequence of each fragment determined) suffer from the inability to detect and quantify any but the most abundant proteins within the sample mixture. For many biological specimens, where the analytes of interest may vary in concentration over 6 orders of magnitude, the current approaches are essentially useless. The number of protein fragments that must be analysed from a human serum specimen in order to sample more than 1% of the constituent proteome is so large as to be impractical. Even the introduction of pre-preparation steps, where the most abundant proteins of all, such as serum albumin, are selectively removed prior to analysis only slightly improve the performance. In principle, such approaches are unlikely ever to provide a rich sampling of the low- and mid-abundance components of the proteome.

Another whole-sample approach is the use of protein-chip (microarray) technology. The principle here is identical to gene chips genomics (which detects the interaction of DNA or RNA in the test sample with a DNA probe on the chip surface). Instead of DNA probes, antibody molecules are coated onto the microarray and the binding of the antigen to the antibody can be quantitated. Such approaches avoid the limitations of other whole sample approaches: like DMI, they can in principle quantitate proteins irrespective of their relative abundance in the test sample. Unfortunately, this approach has a number of limitations – most severe is the inherent lack of quantitative robustness in the microarray detection methodology. The same limitations which reduce the repeatability in micro-array based genomics also prevent the widespread adoption of micro-array based proteomics.

Consequently, there is a need for new proteomic technology which combines all the desirable characteristics of such a technology: it should be a rapid, high throughput approach which avoids the use of technically specialised procedures or capital intensive equipment, and which provides an unbiased sampling of the proteome

irrespective of the absolute abundance of the components present, and which is quantitatively robust under routine laboratory conditions.

Summary of the Invention

5

The present invention provides methods which allow the relative concentrations of many proteins in a pair of samples to be rapidly determined. A tagged antibody library is exposed to a mixture of the test sample and the reference sample, where the reference sample has been labelled in some way. For a given antibody, the amount of label that is bound will be inversely proportional to the amount of the cognate antigen present in the test sample. The amount of label bound to each tagged antibody is read in turn to generate a vector describing the relative pattern of protein concentrations in the two samples.

10

Accordingly, the present invention provides a method of determining the relative abundance of a plurality of proteins in a test sample compared to a reference sample, the method comprising (a) providing a reference sample comprising a plurality of labelled proteins, (b) incubating a plurality of tagged antibodies capable of binding components of the reference sample with (i) a mixture of the labelled reference sample and the test sample and (ii) the reference sample alone, under conditions suitable for the binding of said antibodies to their targets, (c) comparing the amount of labelled protein bound to individual antibody tags in the presence and absence of the test sample.

20

Methods falling under this embodiment may be useful for proteomics (the science of studying large populations of proteins simultaneously). An example of such a proteomic application would be in clinical diagnostics, whereby measuring the levels of many proteins in a biological specimen simultaneously could be used to make a diagnosis of a disease or condition.

25
30

The same principle may also be applied to the profiling of the array of antibodies that are present in a sample, for example the array of antibodies made by different

individuals. Such a profile may be diagnostic of the immune status of the individuals from whom the samples were obtained.

The present invention also provides a method of detecting a plurality of immunoglobulins in a test sample, the method comprising (a) providing a plurality of tagged antigens, (b) incubating said tagged antigens of (a) with said test sample, under conditions suitable for the binding of any immunoglobulins present in said test sample to their targets, (c) incubating said mixture of (b) with one or more labelled antibodies capable of binding specifically to immunoglobulins, (d) measuring the amount of labelled antibody bound to each tagged antigen.

In a further aspect, the invention provides a method of reducing the redundancy and bias of an antibody-expressing phage library comprising:

- (a) providing two surfaces to which a sample of antigens is bound wherein said antigens are bound to the second surface at a higher density than to the first surface;
 - (b) exposing a phage display library to a first surface of (a) under conditions suitable for antibody binding and selecting phage bound to said surface;
 - (c) exposing said selected phage of (b) to a second surface of (a) under conditions suitable for antibody binding and selecting phage not bound to said surface;
 - (d) optionally further selecting said phage of (c) according to steps (b) and (c) one or more times;
- thereby obtaining a library of antibody-expressing phage which has reduced redundancy and/or bias characteristics compared with the original library. An antibody library obtained by such a method may be tagged and used in a screening method of the invention.

Brief Description of the Figures

Figure 1: Schematic representation of two embodiments of the invention.

A: A library of antibodies against the proteins of interest is constructed. Such a library should be highly representative of the proteins in the sample under test, and have a low degree of redundancy (so that antibodies against the same protein do not occur more than a small number of times in total in the whole library). This library is then tagged using one of a range of commercially available tagging technologies, such as the SmartBead platform that uses aluminium barcode tags made by semiconductor fabrication technology.

The specimen under test is then mixed with a reference specimen which has been labelled with a suitable label (for example a fluorescent marker). The mixture of test and reference samples is then incubated with the tagged antibody library and the amount of labelled protein that binds to its cognate antibody is influenced by the amount of the same protein present in the unlabelled test sample. If the protein level is higher in the test sample, the amount of label bound to the tagged antibody is decreased, while if the protein level is lower in the test sample, the amount of label bound to the tagged antibody is increased.

The library is then passed through a laboratory flow cytometer that can read both the tag and barcode and quantify the amount of fluorescence label bound. This approach may be capable of generating up to 1 million datapoints in 15 minutes. Provided that the redundancy of the antibody library is very low, this translates into a relative measure of the level of hundreds of thousands of proteins.

The protein profile that is generated (a vector containing many numbers representing the relative levels of fluorescence bound to each of the tagged antibodies) can be analysed by conventional megavariate pattern recognition methods and provide a protein "fingerprint" for the sample class under study.

B: An antigen library is generated and coupled to the tags, analogous to those in A. This library is then exposed to the test sample of human serum and antibodies in the serum bind to the library of antigens. Any bound human immunoglobulin is then detected by addition of a standardised solution of anti-Ig antibodies labelled with

different fluorophores. For example, by using anti-IgG labelled with the green fluorophore fluorescein and anti-IgM labelled with the red fluorophore rhodamine it is possible to simultaneously quantify the amount of each immunoglobulin subclass which binds to each antigen in turn.

5

Figure 2: A chromatogram of a typical reference sample after labelling the protein with fluorescein isothiocyanate, as described in the text. The labelled sample is applied to a Sephadex G25 column and the eluate is monitored at 280nm (A280) and 450nm (A450). The labelled protein elutes first (around 10-20ml) and has high A280 and A450. The free label elutes much later in a broad peak and has much higher A450 than A480.

10

Figure 3: A graphical representation of the DMI-derived proteomic profile of Individual A, based on data taken from Table 2. The height of the bar from the origin represents the percentage of the population variance exhibited by this individual. The depth of colour represents the absolute deviation of the signal from 1 arbitrary unit. Large, deep coloured boxes contain the majority of diagnostic information about the individual.

15

Figure 4: Impact of iterative rounds of positive selection (at low protein density on the selection surface) followed by negative selection (at high protein density on the selection surface) on the bias of a phage library. Bias was calculated by direct ELISA for phage binding to serum albumin (A) or Fibrinogen (B) or PAI-1 (C) or TGF- β (D) according to the formula $(A+B)/(C+D)$, expressing the direct ELISA result as fraction in the range 0 to 1 representing the total phage concentration required to obtain a half-maximal signal. Error bars are SEDs calculated by assuming A and B to be estimates of the same parameter and C and D to be estimates of the same parameter. Four rounds of this selection protocol reduced the bias factor of this library by approximately 8 fold.

25

30

Definitions

"(Library) component": A single antibody, protein or other antigen, or a mixture of antibodies, proteins or antigens, that are attached to a uniquely coded pool of tags. There may be many individual tags composing such a component, but they will all have the same code. Similarly, there may be many molecules of the antibody,
5 protein or antigen but they will be identical, or else all come from the same mixture.

"Master Library": A library of components which is much larger and more complex than a DMI library. A DMI library can be generated by sub-selecting just a fraction of the components from a master library. Typically such a master library will be
10 composed of more than 10 million components.

"DMI Library": A library made up of components which is suitable for DMI. Typically, such a library will be composed of between 10 and 1 million components, more typically between 100 and 10,000 components.

15

"Tag": Any method of rapidly and easily determining the identity of an antibody, protein or other antigen bearing the tag. Tags are distinguished from "Labels" (see below) by their categorical property: that is, tags need only contain nominal information (tag 1, tag 2, tag 3 and so forth) and not necessarily any continuous
20 information (a variable ranging from 0 to infinity).

"Label": Any method of rapidly and easily determining the amount of an antibody, protein or other antigen bearing the label. Labels are distinguished from "Tags" (see above) by their quantitative property: that is, labels need only contain continuous
25 information (a variable ranging from 0 to infinity) and not necessarily any nominal information (label 1, label 2, label 3 and so forth).

"Specific Binding": An antibody specifically binds to a protein or antigen when it binds with high affinity to the protein or antigen for which it is specific but does not
30 bind, or binds only with low affinity, to other proteins. For example, the antibody may bind to the protein or antigen with 5 times, 10, 20 times, more affinity than to a randomly generated polypeptide or other molecule.

Detailed Description of the Invention

The method of the invention is generally termed "Differential Megaplex Immunoassay" technology (DMI) herein. This strategy provides a relative abundance for each protein component in the proteome, compared to a reference sample (hence the term "differential"). It allows the analysis of thousands or even millions of proteins simultaneously (hence the term "megaplex", which is a higher order extension of the conventional term multiplex). The key analytic technique exploited is the competition immunoassay (hence the term "immunoassay").

1. DMI for Proteomic Profiling

In general terms, to perform a DMI experiment for proteomic profiling you require: an antibody library, a method of tagging the antibodies so that they can be uniquely identified, a reference sample, a method of labelling the reference sample and a strategy for reading the amount of label bound to each tagged antibody. Any or all of the components of the DMI experiment may be already known in the public domain, but the principle of combining these techniques in order to perform proteomic analysis is novel, and represents the invention described herein.

The general principle of the DMI experiment is as follows (see Figure 1A):

1. Mix the labelled DMI reference sample with the sample under test, preferably in equal proportions;
2. Add the tagged antibody library and incubate together;
3. Read the amount of label bound to each tagged antibody.

First, the requirements for each of the key components of the experiment are described, followed by an exemplification of the general DMI experiment laid out above.

A: The antibody library

To be useful for DMI, the antibody library to be utilised should contain a significant number of antibodies which have as their cognate epitopes proteins that are present in the sample to be analysed. For example, to perform a proteomic screen using DMI on a human serum sample would require a library of antibodies a significant proportion of which recognised proteins present in human serum samples.

Ideally, such a library will also have a high degree of complexity: that is, that most, if not all, of the individual antibody species that compose the library, should recognise different proteins. In one embodiment, therefore, each of the plurality of antibodies used in the methods of the invention recognises and binds a different protein. Each antibody may recognise and specifically bind a different protein. Libraries with a high degree of redundancy, by contrast (where many of the antibody components recognise the same protein), will reduce the power of the DMI approach.

Ideally, the library should contain a large number of antibodies. An antibody library useful for DMI may contain between ten and 100 million antibodies, more typically between one hundred and 1 million antibodies.

The library must exist in a format where by the antibodies against different proteins are physically separated, or capable of physical separation. This ensures that each individual antibody component of the library can be uniquely tagged.

Antibody libraries with these properties can be constructed in a number of ways. For example, antibodies known to recognise components of the proteome of the sample to be investigated could be purchased individually from commercial antibody sellers, or else manufactured individually by the standard methods well known in the art. Libraries compiled in such a way are likely to be at the lower end of the size useful for DMI (typically 100 or less antibodies).

Alternatively, the library may be generated by phage display technology. A sample typical of those to be subsequently analysed by DMI may be coated onto a surface and used to positively select antibodies from very large general purpose libraries

(such as those owned and generated by Cambridge Antibody Technology Limited, and similar companies). An antibody library generated in this way may, however, not comply with the ideal characteristics of a DMI antibody library in several ways - the redundancy may be relatively high and the population may be biased by the amount of each protein present in the positive selection mixture.

The present invention therefore provides a modification to the procedure well known in the art for selecting from phage display libraries which allow a low redundancy library with relatively little bias on amount of antigen present to be developed:

In order to reduce the bias of the library towards abundant species in the selection mixture, rounds of positive and negative selection are repeated iteratively, adjusting the total protein concentration applied to the selection surface. In the first round of positive selection, the selection mixture is applied at very low total protein concentration, for example from $0.1\mu\text{g}$ to $100\mu\text{g}$ per cm^2 , to a very large surface area. This ensures that every protein the sample is efficiently represented on the surface. Phage are positively selected, released and grown up back up in number. This selected population is then subjected to a round of negative selection, where the same selection mixture as used in the first round is now applied to the surface at very high total protein concentration, for example 1 mg per cm^2 upwards, over a very small surface area. As a result, many of the phage directed against the abundant antigens bind to the surface and are lost from the population, whereas stochastically the rare proteins will hardly be represented on the negative selection surface where surface area for protein binding was limiting. The population of phage in the supernatant after negative selection are again grown up, and the process can be repeated iteratively with alternate round of positive selection and negative selection.

Preferably the high protein density selection is carried out at a protein density between 10 and 10,000 fold higher than the low protein density selection, more preferably between 100 and 1,000 times higher density. These ranges are based on the use of commercially available high-protein capacity plastic surfaces currently available (such as Nunclon plastics used to make ELISA plate wells) but may need to

be adjusted accordingly for other substrates with different total protein binding capacities. Typically, the low protein density selection should be performed between 100 and 1-fold lower density than the nominal protein binding capacity of the substrate, preferably about 10-fold lower. The high protein density selection should
5 be performed between 1-fold and 100-fold higher density than the nominal protein binding capacity of the substrate, preferably about 10-fold higher. The higher the high protein density coating concentration is relative to the nominal protein binding capacity of the substrate, the more extreme will be the change in library bias.

10 The bias of the library may be assessed as follows: the number of individual library components which bind to two different proteome components which are known to be highly abundant in the samples of interest (in the case of serum, these might be albumin and fibrinogen, for example) are determined. Similarly, the number of library components binding to two rare proteome components are also determined
15 (cytokines such as TGF-beta and MCP-1 would be suitable markers for human serum). Direct ELISA may be used to quantitate the fraction of the total library elements that bind to each of these four marker proteins. The bias of the library would be calculated as $(A + B) / (C + D)$ where A and B are the number of library elements binding to the abundant protein markers, and C and D are the number of
20 library elements binding to the rare protein markers. Initially, after the first round of positive selection, this Bias Factor may be 1,000 or more. After several iterative rounds, the Bias Factor will approach 1.

The Bias Factor of the resulting library may decline faster if the ratio of the protein
25 density on the selection surface during positive selection to the protein density on the selection surface during negative selection is stepwise reduced as the number of selection rounds is iterated. An example of such a selection protocol is illustrated in Figure 4.

30 A DMI Antibody Library generated by phage display approaches will likely contain 10,000 to 10 million distinct antibody components and will, therefore, likely be at the upper end of library size useful for DMI.

To allow for unique tagging of each antibody component, the DMI antibody library may need to be formatted in a manner that physically separates the library components. For libraries where each component is generated individually, the components could be dispensed one at a time into multiwell plates, for example, at a known antibody concentration. For libraries generated by phage display approaches, multiple individual phage clones could be grown up, for example in multiwell plates, and the antibody concentration normalised in each well.

10 B: Method for tagging the antibody library

DMI requires that each antibody component of the library be uniquely tagged in a manner that allows the antibody to be identified when in a mixture. Any method of tagging which allows the antibody to be identified, while still retaining its ability to specifically bind to its antigen, would be suitable for use in DMI.

15

Examples of suitable tagging methodologies would include:

Aluminium bar codes (such as those developed by Sentec Ltd). These bar codes are 100µm x 10µm x 1µm aluminium strips which have holes punched in them, allowing millions of unique codes to be stamped onto them. They are produced using semiconductor chip fabrication methodology to very high specification. Each tag code is handled separately, for example in different wells of multiwell plates. The tag and the antibody can be coupled together by any method obvious to those skilled in the art, including heterobifunctional crosslinking or by charge-coatings applied to the tag. Any method that irreversibly couples the tag to the antibody without denaturing the antibody would suffice.

Dye-impregnated beads (such as those developed by Luminex). The beads have dyes with unique spectral properties impregnated into them, which can be used to unambiguously identify the bead. Dye-bead technology would likely only be useful for smaller DMI antibody libraries (less than approximately 100 antibody components) because of the limited availability of enough different suitable dyes.

The bead and the antibody could be coupled together by any method obvious to those skilled in the art, including heterobifunctional crosslinking or by charge coatings applied to the bead.

- 5 Each tag may be linked to one or more antibody species. In one embodiment, each antibody species within the library is linked to a different tag so that the binding of each antibody may be assessed separately. Alternatively, two or more antibody species may be linked to a tag. For example, different antibody species which bind the same or different epitopes in a target protein may be pooled and linked to a single tag. In this way, all antibody binding to that target protein may be determined by assessing the label associated with that tag.

Irrespective of the tagging technology used, the ratio of antibodies per tag could be controlled, depending on the coupling chemistry selected. For DMI applications it would be desirable to have a large number of antibody molecules attached to one tag (from 10^{11} to 10^{15} or more antibody molecules per tag) since the signal to noise ratio for reading the bound label will increase with increasing antibody density on the tag.

C: The Reference Sample

- 20 DMI is a differential assay methodology: it does not measure the absolute level of any analyte within the test sample, but estimates the ratio of the amount of the analyte in the test sample compared to a reference. Consequently, each DMI experiment requires a reference sample. The reference sample should be the same for every DMI experiment where the resulting protein profile data are to be compared.

The reference sample should be of similar overall composition to the test samples - it should contain the same analytes in approximately the same concentrations as the test sample. For example, a reference sample may be obtained from the same tissue as the test samples. A reference sample may be obtained from the same species as the test samples. Preferably, the reference sample is obtained from the same tissue in the same species as the test samples. DMI shows excellent quantitative resolution where

the ratio of the analyte is close to 1 (say, in the range 0.1 to 10) but outside these ranges the signal gradient declines sharply. Consequently, to obtain the highest data density in the resulting protein profile, the concentration of each analyte in the reference sample would ideally be equal to the average of the analyte concentration in all the test samples.

One method of generating such a reference sample would be to take a small amount of all the samples to be tested and pool them, mixing thoroughly. The resulting pool would have the ideal properties of a reference sample for DMI.

Another method for generating a reference sample would be to make a pool of samples of similar origin to the test samples, but not actually including the test samples. The use of pooled reference samples increases the likelihood that: (a) every analyte present in the test sample will be represented in the reference sample and (b) that the concentration of each analyte in the reference sample approaches the average value for all the test samples.

As an example, to create a reference sample for a DMI experiment examining human serum samples; aliquots of serum from many different human subjects may be taken and pooled. To create a reference sample for a DMI experiment examining cultured liver cells, protein extracts from many different cultures of liver cells would be taken and pooled. It would not be appropriate to use a pool of human liver cell extracts as the reference sample for a DMI experiment examining human serum samples.

After labelling (see below), the reference sample should be at approximately the same total protein concentration as the average of the test samples. If necessary, the total protein concentration of the labelled reference sample should be adjusted prior to beginning the DMI experiment.

D: A method for labelling the reference sample

The reference sample is labelled such that a plurality of proteins within the sample bear the label. In a preferred embodiment, the reference sample is labelled in such a

fashion that all of the protein components within the sample are labelled to some extent. Each different protein component may or may not be labelled to the same extent as all the others.

5 Any label may be used which can be read easily and rapidly once bound to the tagged antibodies. For example, the label may be a fluorescent dye that can be read by interrogating the tagged antibody with a laser, inducing fluorescence, which can be quantitated with a photodetector.

10 Suitable fluorescent dyes include: fluorescein, oregon green, GFP, rhodamine, r-Phycoerythrin, Cy3, Cy5, coumarin, AMCA, texas red, Alexa Fluor dye series (350, 430, 488, 532, 546, 555, 568, 594 and 633) and BODIPY series (493/503, FL, R6G, 530/550, TMR, 558/568, 564/570, 576/589, 581/591, TR, 630/650-X and 650-655-X). Providing appropriate post-processing steps are utilised (which are well known
15 in the art) then lanthanide chelates can be used as labels (for example Europium chelates) which are read using laser-induced fluorescence which has a very long lifetime, allowing time-resolved fluorescence reading to improve signal to noise ratios. Alternatively, a non-fluorescent label could be used. Suitable non-fluorescent labels include: radioactive decay (for example: tritium, iodine-125, phosphorus-32,
20 sulphur-35 labels; read using a suitable scintillation counter), gold particles of various sizes (read using a microscope, preferably with automated image analysis software to identify and count the particles) and chemiluminescent probes (for example luciferase label read by exposing it to luminol-containing buffer in a
luminometer).

25

The chemistry used to couple the label to the protein components of the reference sample must meet three criteria: (a) it must irreversibly couple the label to the protein (b) the protein must not be denatured by the process and (c) the label must still be detectable after the coupling reaction. Any chemistry that meets these criteria can be
30 used. For example, fluorescein isothiocyanate can be reacted with the protein fraction of the reference sample. After removal of unconjugated fluorescein e.g. by

column chromatography) the labelled sample can be reconstituted to a total protein concentration equal to the approximate average of the test samples.

The labelling ratio (the number of labels per protein molecule) can vary within a reasonable range for a DMI reference sample. Typically it will be in the range 0.1 to 50 labels per protein, more typically in the range 1 to 5. Low labelling ratios reduce the sensitivity of the detection system, and increase noise, while high labelling ratios can affect the ability of the labelled protein to bind to its cognate antibody in the tagged antibody library.

E: Strategy for reading the amount of label bound to each tag

The strategy for reading the amount of label bound to each tag will depend on the nature of the tag and the label. In order to generate data-rich protein profiles the reading method should be relatively high throughput. However, for small DMI antibody libraries (e.g. less than a few hundred antibody components) the label could be read manually. For example, using a microscope each tagged antibody in turn could be identified and the tag read, then the amount of label determined. Reading the tag might involve, for example, taking a spectrum of the tagging dye or reading the aluminium bar code under transmission illumination. Reading the label might involve, for example, counting bound gold particles or capturing induced fluorescence with a photomultiplier.

For larger DMI antibody libraries (with thousands or millions of antibody components) an automated strategy for reading each tagged antibody component will be required. For example, the tagged antibody components could be passed one at a time through a standard flow cytometer. In the example where the tag is an aluminium bar code and the label is a fluorescent dye, the flow cytometer (with appropriate software) could read both the tag and the bound label.

Successful DMI requires that both the reading of the tag and the bound label be performed with high fidelity and reproducibility. For example, for the determination of bound label on a bar-code tagged antibody, a standard flow cytometer can read the

tag correctly with an error rate of less than 1 in 10,000, while the estimate of bound fluorescent label can be performed with a repeated measures coefficient of variation below 5%. With these characteristics, DMI approaches the robustness of methods such as NMR-based metabonomics, while retaining the ease, speed and cost benefits of gene array technology.

F: The procedure

The labelled reference sample, adjusted to the same total protein concentration as the average of the test samples, is then dispensed at an appropriate volume into tubes or microtitre plate wells. Typically volumes between 1µl and 200µl will be used.

Next, each test sample is added one well at a time. The volume of test sample is preferably equal to that of the labelled reference sample. The plate must then be mixed thoroughly, to ensure the test and reference samples are homogeneously distributed.

An appropriate volume of the mixed antibody library must then be added. Typically between 1µl and 100µl of library will be added. The number of individual tags to be added will depend on the complexity of the library, as well as its redundancy and bias factors. Typically, between 10 and 200 times more individual tags will be added than there are non-redundant components of the library. After addition of the library, the reaction tubes or plates must be mixed thoroughly, and incubated under conditions suitable for the binding of the antibodies to their targets, for example for a period to allow the antigens in the test and reference samples to bind to their cognate tagged antibodies. Typically, this will be for a period between 10 and 180 minutes. Typically, the reactions will be continually agitated throughout the incubation to ensure that the tags remain randomly suspended within the liquid. Typically, the incubation will be performed between 4°C and 37°C. Other components may be added to the reaction as appropriate, to improve the specificity and selectivity of

antibody binding to antigen: typically, a non-ionic detergent is added at a concentration between 0% and 1% volume/volume (for example, Tween 20 at 0.1% v/v). Similarly, the salt concentration can be varied: typically, sodium chloride

solution is added to increase the total salt concentration by between 0mM and 250mM. Similarly, the divalent cation concentration can be varied: typically, calcium chloride or magnesium chloride are added to increase the calcium or magnesium ion concentration by between 0mM and 10mM as required, or EGTA is added to decrease the calcium and magnesium concentrations as required. Similarly, the pH of the reaction can be varied: typically, 1M hydrochloric acid or 1M sodium hydroxide are added to reduce or increase, respectively, the pH of the reaction by between 0 and 3 units.

At the end of the reaction, the interaction between antigen and antibody is typically terminated. Several methods can be used: for example, the reactions can be diluted substantially (typically by 5 to 50 fold with buffered saline); alternatively, the reaction can be rapidly cooled (typically to 4°C); alternatively a crosslinking reagent can be added (typically formalin is added to a 3% final concentration).

Following termination of the reaction, the tagged antibodies can be directly read or they can be washed by gentle ultrafiltration and then resuspended at an appropriate concentration prior to reading. Whether the tagged antibodies need to be washed prior to tagging will depend on the method of reading. Typically, using a fluorescence microscope or a flow cytometer, no washing step is necessary.

The amount of label bound to each tag must then be determined. The number of tags which must be read varies depending on the complexity of the library, as well as its redundancy and bias. Typically, between 2 and 200 tags will be read for each non-redundant component of the library. The smaller the library, the larger the number of tags per component that can be read. If low numbers of tags per component are read for very large libraries, then a significant number of components in the final vector will have to be recorded as data missing values. Where more than one tag representing the same component is read, the amount of label bound to each is typically averaged before reporting the final vector.

The resulting output vector can then be analysed in a number of ways. Typically, a number of vectors from different individuals are used to construct the X-matrix for various megavariable statistical analyses, including PCA, PLS-DA and OSC. Such methods allow the individuals to be classified according to some pre-existing phenotype (such as disease status). Once a model has been constructed classifying individuals whose phenotypic status is known, the model can then be used to predict the phenotype of individuals whose status is unknown. This is the basis of the application of DMI proteomic profiling to medical diagnostics.

- 10 The DMI approach has a number of advantages over current proteomics platforms. In particular, existing methods can be limited in sensitivity to the relatively abundant components in the mixture. For example, when applied to serum, the very high levels of albumin in the sample can hamper traditional approaches. However, provided that the antibody against albumin is present only once in the tagged DMI library then
- 15 albumin will contribute only one data point to be protein profile. DMI is also quantitatively robust, with coefficients of variation below 5% for most antibodies, and therefore substantially superior to microarray-based proteomic platforms.

2. DMI for Immunomics

- 20 Immunomics is a newly coined term for a highly specialised example of proteomics: analysis of the population of antibody molecules produced by a given individual at a given time. This information is not normally encoded within a proteomic profile (whether generated by DMI or classical methods). It is also absent from genomic, transcriptomic or metabonomic datasets. Consequently, specialised techniques will
- 25 be required to perform high throughput analysis of the immunomic repertoire. To date, there are no publicly disclosed methods for performing immunomics. Consequently, a second important application of the DMI principle is as a first high throughput, robust and reproducible method for obtaining an immunomic dataset.
- 30 In general terms, to perform a DMI experiment for immunomics you require: an antigen library, a method of tagging the antigens so that they can be uniquely identified, one or more labelled anti-immunoglobulin antibodies and a strategy for

reading the amount of label bound to each tagged antibody. Any or all of the components of the DMI experiment may be already known in the public domain, but the principle of combining these techniques in order to perform immunomic analysis is novel, and represents the invention described herein.

5

The general principle of the DMI experiment is as follows:

1. Mix the tagged antigen library with a test sample;
2. Detect bound antibody with a panel of labelled anti-immunoglobulin antibodies;
- 10 3. Read the amount of label bound to each tagged antibody.

First, the requirements for each of the key components of the experiment are described, followed by an exemplification of the general DMI experiment laid out above.

15

A: The antigen library

The requirements for the antigen library for immunomics are very similar to the requirements for the antibody library for proteomic profiling: the library should be as large as possible with low redundancy (preferably with any given antigen only
20 represented by a single component of the library).

A suitable antigen library may comprise oligopeptides and/or oligosaccharides. The source of the antigens can either be by manual assembly of the library using purified protein and non-protein antigens as individual library components (analogous to the
25 manual assembly of an antibody library using purified antibodies) or generated by combinatorial chemistry. For example, a peptide antigen library could be generated by standard solid phase chemistry, using methods well known in the art.

As with the antibody library, the components of the antigen library must be capable
30 of being separated (or else be generated separately) so that they can be dispensed individually (for example, into microtitre plates) to allow them to be tagged.

B: A method of tagging the antigen library

All of the same considerations that applied when tagging the antibody library described above apply to tagging the antigen library, and the same methods are likely to be useful. Where the library components are proteinaceous, then the antigen library can be treated exactly as if it was an antibody library. Where the library is composed of oligopeptides, then consideration of the tagging can be incorporated into the synthetic chemistry used to generate the antigen: for example, a chemical linker can be added to every peptide during synthesis, and this linker can be used to attach the peptides to the tags. The precise nature of the linker would vary depending on the nature of the tag. For dye-containing latex beads, for example, a bifunctional succinamide crosslinker could be utilised. Where the library is composed of oligosacharrides, then the sugar chains can be attached to a carrier protein and then the library be treated as for a protein library, or else a suitable crosslinker can be added to the sugar chains during synthesis, as for the peptides.

C: A panel of anti-immunoglobulins appropriately labelled

Whereas, for proteomic profiling the label is applied to the reference sample, and the amount of each protein in the test sample is measured indirectly by competition with the labelled reference sample, for immunomics the antibody that binds to each tagged antigen is directly detected. This requires a panel of anti-immunoglobulins, or equivalent reagents, which bind to immunoglobulins with high affinity and specificity.

The anti-immunoglobuline should be specific to the types of immunoglobulin likely to be present in the test sample. For example, the anti-immunoglobulins may be specific to immunoglobulins from the same species as the test sample, e.g. anti-human immunoglobulins where the sample is derived from a human.

Suitable immunoglobulin panels are readily available from commercial sources - for example, the WHO standard antibodies for detecting human immunoglobulins can be used. In the ideal experiment, a panel of one or more such antibodies would be used as detection reagents, one specific for each of the heavy chain classes of

immunoglobulin found in the required species. For example, a panel of antibodies specific to one or more of the heavy chain subclasses in humans (IgG1, IgG2a, IgG2b, IgG3, IgG4, IgA, IgD, IgE and IgM) may be used. The WHO standard antibodies are mouse monoclonal antibodies, and are consequently available in large, and essentially inexhaustible batches of detection reagents with identical properties.

The selected detection reagents must then be labelled using any method suitable for high throughput detection as described above in relation to the labelling of the reference sample in proteomics. For example, the WHO standard antibodies can be labelled with fluorescent dyes. A different dye may be used for each different detection reagent (for example, anti-human IgG1 could be labelled with fluorescein, while the anti-human IgM could be labelled with r-Phycoerythrin). There are plenty of spectrally distinguishable fluorescent dyes available to allow all nine of the WHO standard antibodies to be separately quantitated.

As for the labelling of the reference sample for protein profiling, the only other requirement for the label is that it does not affect the detection characteristics of the detection reagent once the label is applied, and that the label can still be read once it has been bound to the detection reagent. The same requirement applies here.

D: A strategy for reading label bound to the tagged antigen library

All of the considerations that applied to reading a tagged antibody library for DMI proteomic profiling, also apply identically to reading a tagged antigen library for DMI immunomic profiling.

E: The procedure

The test samples, e.g. serum samples are added one well at a time, dispensing an appropriate volume of each (typically 1µl to 200µl).

An appropriate volume of the mixed antigen library is then added. Typically between 1µl and 100µl of library will be added. The number of individual tags to be added will depend on the complexity of the library. Typically, between 10 and 200

times more individual tags will be added than there are components of the library. After addition of the library, the reaction tubes or plates must be mixed thoroughly, and incubated under conditions suitable for the binding of any antibodies present in the test sample to their targets, for example for a period to allow the antibodies in the test serum to bind to their cognate tagged antigens. Typically, this will be for a period between 10 and 180 minutes. Typically, the reactions will be continually agitated throughout the incubation to ensure that the tags remain randomly suspended within the liquid. Typically, the incubation will be performed between 4°C and 37°C. Other components may be added to the reaction as appropriate, to improve the specificity and selectivity of antibody binding to antigen: typically, a non-ionic detergent is added at a concentration between 0% and 1% volume/volume (for example, Tween 20 at 0.1% v/v). Similarly, the salt concentration can be varied: typically, sodium chloride solution is added to increase the total salt concentration by between 0mM and 250mM. Similarly, the divalent cation concentration can be varied: typically, calcium chloride or magnesium chloride are added to increase the calcium or magnesium ion concentration by between 0mM and 10mM as required, or EGTA is added to decrease the calcium and magnesium concentrations as required. Similarly, the pH of the reaction can be varied: typically, 1M hydrochloric acid or 1M sodium hydroxide are added to reduce or increase, respectively, the pH of the reaction by between 0 and 3 units.

At the end of the reaction, the tags are washed by gentle ultrafiltration, typically with phosphate buffered saline. Other components, such as non-ionic detergent can be added to the wash buffer to improve the specificity and selectivity of antibody binding to antigen. Typically, Tween 20 is added at 0% to 1% volume/volume final concentration.

After washing, the tags are resuspended in a buffer containing the panel of labelled detection reagents. For example, where the test sample is from a human source, anti-human immunoglobulin antibodies are used as detection reagents at a concentration between 0.05 and 50µg/ml for each individual antibody (more typically between 0.5 and 5µg/ml). Additional components can be added to the incubation buffer to

improve the specificity of detection reagent binding to the captured antibody on the tags. These are the same components that could be added during the initial reaction of the library with the test samples. The labelled detection reagents are then typically incubated with the tagged library for between 10 and 180 minutes. The reactions are typically agitated for the period of the incubation to keep the tags randomly suspended in the liquid. The incubation is typically performed at between 4°C and 37°C.

At the end of the reaction, the tags may be washed by gentle ultrafiltration, typically with phosphate-buffered saline. Other components, such as non-ionic detergent can be added to the wash buffer to improve the specificity and selectivity of antibody binding to antigen. Typically, Tween 20 is added at 0% to 1% volume/volume final concentration. Whether the tagged antibodies need to be washed prior to tagging will depend on the method of reading. Typically, using a fluorescence microscope or a flow cytometer, no washing step is necessary.

The amount of label bound to each tag must then be determined. The number of tags which must be read varies depending on the complexity of the library, as well as its redundancy and bias. Typically, between 2 and 200 tags will be read for each non-redundant component of the library. The smaller the library, the larger the number of tags per component that can be read. For each tag, the amount of each different label (representing each of the different heavy-chain classes of immunoglobulin) will be read separately. Depending on how many immunoglobulin classes were separately detected, the output vector will have between one and nine times more values than there are non-redundant components to the library. If low numbers of tags per component are read for very large libraries, then a significant number of components in the final vector will have to be recorded as data missing values. Where more than one tag representing the same component is read, the amount of label bound to each is typically averaged before reporting the final vector.

30

The resulting output vector can then be analysed in a number of ways. Typically, a number of vectors from different individuals are used to construct the X-matrix for

various megavariable statistical analyses, including PCA, PLS-DA and OSC. Such methods allow the individuals to be classified according to some pre-existing phenotype (such as disease status). Once a model has been constructed classifying individuals whose phenotypic status is known, the model can then be used to predict the phenotype of individuals whose status is unknown. This is the basis of the application of DMI proteomic profiling to medical diagnostics.

Examples

10 Example 1 : A proteomic analysis of human serum using a small antibody library, aluminium bar-code tags and a fluorescein labelled reference sample

In the first step, an antibody library suitable for use in DMI was generated. For this pilot demonstration of the invention, the library was constructed by obtaining quantities of purified antibodies against human serum components from a range of manufacturers. Each of the antigens to be studied was included in the library just once, and as a result the library had the ideal characteristic for DMI libraries of very low redundancy.

20 For this experiment, thirty eight different antibodies were selected. Thirty-four were against distinct serum components (see Table 1). The remaining 4 were control antibodies of the same species as the 34 antibodies, but with epitopes selected to be absent from the reference sample. The 34 serum components to be detected in this experiment ranged in abundance from albumin (~30mg/ml) to IL-1b (100pg/ml).
25 However, for three of the antibodies against the least abundant components (anti-HIVp24gag, anti-soluble selectin and anti-IL1b) no signal was detected in the reference sample and consequently no data was obtained from these tags. The least abundant protein to be robustly detected in our experiment was TGF-beta (~30ng/ml), representing a working dynamic range for DMI of approximately 1
30 million fold. Since each antibody was purchased separately, they were available in 38 separate containers, allowing them to be dispensed at an antibody concentration of 1mg/ml in phosphate-buffered saline into wells of a microtitre plate.

Table 1

Tag	Antigen	Antibody	Species	CVar
1	α 2-macroglobulin	Biogenesis 5850-0004	Sheep IgG	3.8
2	α 1-antitrypsin	Calbiochem 178260	Mouse IgG2a	2.1
3	ApoAI	Calbiochem 178422	Rabbit IgG	7.2
4	ApoB	Calbiochem 178426	Rabbit IgG	11.4
5	ApoE	Biogenesis 0650-2054	Mouse IgG1	6.8
6	β 2-microglobulin	Sigma M7398	Mouse IgG1	2.3
7	CICP	Quidel 1M0622	Rabbit IgG	2.2
8	Fibrinogen	Biogenesis 4440-8004	Sheep IgG	3.0
9	HIV1p24gag	ARP ARP313	Mouse IgG	-
10	ICAM-1	Serotec MCA532	Mouse IgG1	17.6
11	Ig Kappa LC	Bionostics M03010	Mouse IgG1	2.6
12	IgA	Bionostics M26012	Mouse IgG1	2.4
13	IgD	Bionostics M01014	Mouse IgG1	2.9
14	IgE	Bionostics M38041	Mouse IgG1	8.1
15	IGF-1	Serotec MCA520	Mouse IgG1	2.3
16	IL1 β	R&D Systems	Mouse IgG1	-
17	Lp(a)	Immunoscientific	Sheep IgG	4.5
18	MMP9	Chemicon AB805	Rabbit IgG	3.5
19	Myeloperoxidase	NeoRX NR-ML-5	Mouse IgG	2.6
20	Osteopontin	Hoyer 1826-1283	Rabbit IgG	3.3
21	PAI-1 (free)	Progen TC21173	Mouse IgG1	6.9
22	PAI-1 (complex)	Mol Innovations MA14D5	Mouse IgG1	2.5
23	PAI-2	American Diagnostic #3750	Mouse IgG2a	2.7
24	PDGFAA/AB	UBI #06-130	Rabbit IgG	4.6
25	Selectin E/P	R&D Systems BBA1	Mouse IgG1	-
26	Serum Albumin	Calbiochem 126582	Rabbit IgG	3.8
27	SHBG	Biogenesis 8280-0108	Mouse IgG1	2.6
28	TGF- β 1	R&D Systems BDA19	Chicken IgG	5.0
29	TGF-LTBP	R&D Systems Mab39	Mouse IgG	4.7
30	Thrombospondin	Biogenesis 8835-0004	Mouse IgG1	2.3
31	TIMP-2	Biogenesis 9013-2609	Sheep IgG	3.3
32	TPA	American Diagnostic #387	Goat IgG	2.4
33	UPA	Accurate YMPS75	Goat IgG	2.9
34	VWF	Dako A082	Rabbit IgG	4.6
35	Collagen-II	NIHDHSB CII-C1	Mouse IgG	-
36	NR58-3.14.3	Affiniti ARP063/AF	Rabbit IgG	-
37	Salicylate	Cortex CR1041SP	Sheep IgG	-
38	PPAR-alpha	Santa Cruz sc1985	Goat IgG	-

Table 1: The antibodies that were selected to generate the small manual DMI

5 library are shown above. 'Tag' numbers represent the position of the library

component in the output vector (and is not the code of the tag, which is more complex). 'Antigen' represents the known serum component that the antibody binds to. 'Antibody' represents the source of the particular antibody used. 'Species' is the species of the immunoglobulin fraction used. 'Cvar' is the coefficient of variation for reading multiple tags of the same code in the same experiment. The Cvar is not given for HIVp24gag, ICAM-1 or SelectinE/P because these antigens were below the detection limit of the assay in our reference sample.

This small antibody library was then tagged using aluminium barcode tags. The tags were activated to promote non-covalent protein binding, then mixed with the antibodies: a different bar code was mixed with each component of the antibody library. The tags and antibodies were sealed and incubated overnight to allow the bar code tags to become fully coated in antibody molecules. All the tagged antibodies are then pooled into a single tube, and wash them by gentle ultrafiltration with an excess of phosphate-buffered saline, and resuspended at a known tag concentration (e.g. 1 million individual tags per ml).

In the second step, the labelled reference sample was prepared. Approximately 2ml of pooled serum from 15 healthy volunteers was extensively dialysed against 100mM sodium carbonate buffer pH9 (to remove free amino acids that would prevent the reaction between proteins and the fluorescein isothiocyanate (FITC), as well as to adjust the pH to the optimum for FITC labelling). FITC dissolved in DMSO was then added to the dialysed serum at approximately a molar ratio of approximately 10:1 (serum contains 70mg/ml protein of average molecular mass 50,000 Da, which is equivalent to a concentration of ~1.4 mM; therefore FITC is added to a final concentration of 15 mM. To 2ml of serum, we added 200µl of 150mM stock FITC in DMSO).

The labelling reaction was left to run overnight at 4°C with constant mixing. The reaction was then terminated by addition of 1/10th volume (220µl) of 1M glycine pH 7. The excess glycine rapidly reacts with any free FITC remaining and hence terminates the reaction. The resulting protein mixture is then separated from the

unreacted fluorescein:glycine conjugate by column chromatography. A sephadex G25 column (10ml bed volume) was equilibrated in phosphate-buffered saline, then loaded with the labelled serum sample. The protein component rapidly passes through the column and is collected and retained, while the low molecular weight salts (including the fluorescein) pass much more slowly through the column and are discarded. The separation can be monitored by flowing the column eluate through a dual-wavelength spectrophotometric detector set at 280nm (to observe protein) and 490nm (to observe fluorescein). The trace obtained is shown in Figure 2.

10 The labelled protein eluate from the column was then concentrated using a centrifugal ultraconcentrator (Millipore) with a nominal 3kDa cut-off filter membrane until it was reduced in volume to approximately 1ml - half the original volume of pooled serum. The total protein concentration of this sample was then tested using a Coomassie Plus protein assay (Pierce) with serum albumin as the
15 standard. In our experiment, the protein concentration was 121mg/ml representing a recovery of 86% during the labelling and chromatography steps. An appropriate volume of phosphate-buffered saline was then added to return the total protein concentration of the labelled reference sample to that of the original pooled serum. In our experiment, 730µl of buffer was added to return the total protein concentration
20 to 70mg/ml. This procedure prepared 1.73 ml of labelled reference sample, sufficient for approximately 100 separate assays. The same procedure, however, can be used to prepare much larger batches of reference sample.

In the third step, we performed the actual DMI procedure. In a V-bottom microtitre
25 plate, 20µl aliquots of the labelled reference sample were dispensed. Next, 20µl of each test sample was added to each well - the test samples were undiluted human serum samples, including the 15 samples that had been pooled to create the reference sample pool. The plate was sealed and mixed. Next 10µl of the tagged DMI antibody library (containing about 10,000 individual tags - we aim to add between 10
30 and 200 times as many individual tags as there are discrete components to the library to increase the likelihood that at least one of every tag is included in the mixture) was dispensed into each well. The plate was again sealed, mixed and then

incubated at room temperature for 15 minutes with constant agitation. At the end of the experiment, 150 μ l of phosphate buffered saline was added to terminate the reaction by dilution.

5 In the final step, each reaction in turn was passed through a flow cytometer. For large scale DMI experiments, this can be performed using a robotic autosampler, but for this smaller scale pilot experiment, each reaction in turn was transferred to a FACS tube (Becton-Dickinson) and manually sampled. For each tube 5,000 events were captured (representing 5,000 distinct individual tags). As each tag passed
10 through the laser beam, the time profile of the forward-scatter pulse was decoded to give the binary representation of the tag code. Simultaneously, the FL1 pulse height read at 90° to the incident beam, was taken to represent the amount of labelled protein bound to the tagged antibody. Each pair of numbers (tag code, bound label) were recorded for all 5,000 events. Thereafter, the events were grouped by tag code,
15 and the average bound label for each group of identical codes was calculated. The output from this experiment was a vector with 38 values in tag code order for each of the samples analysed. The results are shown in Table 2 and Figure 3. These profiles represent a proteomic profile for each of the individuals tested, and can be used for various investigation or analytical purposes.

20

In this example, we noted that several of the individuals had elevated levels of the proteins bound to tags 8 and 21 (this is represented by the lower values in Table 2, since high levels of a protein in the test sample reduces the amount of labelled protein from the reference sample which binds to the tagged antibody). These tags
25 had antibodies to fibrinogen and PAI-1 respectively. Since these proteins are both known to be positive acute phase reactants (that is, their levels are known to be elevated during infections), we conclude that these individuals are likely to have been suffering from a minor infection, such as the common cold, at the time the blood sample was drawn.

30

We have performed a full analysis of the sources of variation in the data vector obtained (Tables 1 & 2). Firstly, we have assessed the analytical reproducibility of

the method (Cvar(anal)) calculated from the range of fluorescence readings from different tags with the same code in the same experiment. The analytical reproducibility is excellent (below 5% for most tags, superior to individual immunoassays). Furthermore, the Cvar(anal) is unaffected by the abundance of antigen, being similar for albumin and fibrinogen to TGF-beta and PAI-1.

Table 2

	1	2	3	4	5	6	7	8	9	10
	a2M	a1AT	ApoAI	ApoB	ApoE	b2M	CICP	Fib	HIVp24gag	ICAM1
A	1.105	1.118	1.012	1.470	0.574	1.007	1.698	1.000	-	1.588
B	0.906	0.859	0.957	0.428	0.947	0.914	3.601	0.991	-	1.741
C	0.958	0.951	0.974	1.232	1.524	1.207	0.782	1.235	-	0.121
D	1.287	1.078	0.796	10.096	1.635	1.018	1.156	0.961	-	1.722
E	1.003	0.956	0.622	0.847	1.310	0.923	1.243	1.130	-	1.515
F	0.938	0.982	0.946	7.935	0.775	0.856	0.650	1.465	-	1.544
G	0.967	1.006	2.346	0.759	2.016	0.973	0.600	0.754	-	0.568
H	0.952	0.892	0.949	0.446	0.446	0.960	2.079	1.042	-	1.885
I	1.078	0.844	1.079	0.445	1.171	0.964	4.650	1.065	-	0.963
J	1.113	1.004	1.315	0.738	1.147	1.000	0.636	1.297	-	2.209
K	0.898	1.009	0.770	1.332	1.728	1.040	0.623	1.322	-	0.892
L	0.982	1.133	0.760	4.255	0.943	1.086	2.057	1.009	-	2.602
M	0.942	0.896	0.853	1.123	1.272	0.984	1.496	1.155	-	2.387
N	0.998	0.896	1.009	2.610	1.705	1.006	1.412	0.705	-	0.264
O	1.045	1.095	1.018	1.449	1.098	1.003	0.767	1.015	-	2.560
P1	0.923	0.867	1.315	1.732	1.151	0.992	2.203	0.787	-	0.917
P2	1.070	0.976	1.276	1.239	0.998	1.053	2.261	0.721	-	1.052
P3	0.967	0.933	1.529	1.555	1.123	0.938	2.258	0.753	-	0.522
P4	0.983	1.008	1.338	1.717	1.315	0.973	1.738	0.752	-	0.899
P5	0.972	0.954	1.221	1.698	1.067	1.001	2.207	0.702	-	0.611
Ave	1.011	1.104	1.027	2.344	1.219	0.984	1.563	1.229	-	1.512
Cvar(anal)	3.8	2.1	7.2	11.4	6.8	2.3	2.2	3.0	-	17.6
Cvar(rm)	1.665	3.485	1.540	1.670	3.673	1.943	8.239	1.404	-	10.344
Cvar(indiv)	4.571	3.996	30.131	111.085	26.140	3.805	64.576	14.377	-	24.914

	11	12	13	14	15	16	17	18	19	20
	IgKappaLC	IgA	IgD	IgE	IGF1	IL1b	Lp(a)	MMP9	MPO	OPN
A	0.952	1.224	0.007	1.074	0.866	-	0.109	1.019	0.552	1.054
B	1.067	1.005	12.412	0.958	1.027	-	0.146	0.938	1.406	0.951
C	0.906	1.236	0.037	0.899	1.012	-	27.492	0.984	2.505	1.081
D	1.165	0.997	0.365	1.411	1.061	-	2.432	1.072	1.839	0.996
E	0.949	0.882	0.039	1.122	0.986	-	26.119	1.019	0.736	1.017
F	1.085	1.117	1.304	1.366	1.042	-	22.278	1.066	2.551	0.994
G	0.947	0.907	10.272	0.050	0.933	-	0.149	1.013	0.361	0.911
H	1.030	0.952	11.782	1.267	0.882	-	3.427	1.017	1.632	1.045
I	0.940	0.921	1.974	1.280	0.946	-	0.281	1.031	1.480	1.067
J	1.198	1.088	25.121	1.033	0.938	-	0.106	1.100	0.612	0.857
K	1.164	1.026	0.239	0.881	0.994	-	0.349	0.963	0.936	1.033
L	0.973	0.924	0.004	0.812	0.964	-	31.546	1.136	1.604	1.020
M	1.141	1.672	4.080	0.989	1.054	-	27.395	1.111	1.241	1.085
N	1.007	0.889	0.193	0.925	1.118	-	4.536	1.050	0.341	1.157
O	1.274	1.025	0.525	1.410	1.091	-	1.734	1.050	0.576	0.800
P1	1.006	1.068	10.353	1.388	0.992	-	0.112	1.028	0.672	1.056
P2	1.094	1.100	9.768	0.974	0.979	-	0.128	0.978	0.658	0.973
P3	0.935	1.171	11.951	1.228	0.956	-	0.117	1.067	0.662	1.004
P4	1.067	1.005	13.207	1.401	0.938	-	0.141	1.093	0.667	0.966
P5	1.038	1.094	10.966	1.339	0.996	-	0.112	1.029	0.790	1.062
Ave	1.053	1.058	4.557	1.029	1.026	-	9.873	1.081	1.225	1.005
Cvar(anal)	2.6	2.4	2.9	8.1	2.3	-	4.5	3.5	2.6	3.3
Cvar(rm)	3.379	3.109	9.194	5.872	0.239	-	5.721	0.707	5.556	1.156
Cvar(indiv)	4.769	13.728	147.527	18.759	4.854	-	118.443	1.031	50.726	4.767

	21	22	23	24	25	26	27	28	29	30
	PAI1(f)	PAI1(c)	PAI2	PDGF	Selectin	Albumin	SHBG	TGFB1	LTBP	TSP
A	1.396	0.678	1.021	1.163	-	0.871	0.986	1.152	1.208	1.562
B	0.857	0.692	0.990	1.135	-	1.109	1.060	1.230	1.226	1.489
C	0.908	0.999	1.004	0.944	-	1.018	0.986	0.927	0.980	1.167
D	1.480	0.691	0.964	0.576	-	0.853	1.172	0.579	0.533	0.787
E	1.288	0.954	1.004	1.413	-	1.223	1.008	1.206	1.403	1.609
F	1.323	0.510	0.993	0.667	-	0.896	0.889	0.592	0.621	1.035
G	0.478	0.370	1.034	0.646	-	0.713	1.042	0.638	0.622	0.348
H	1.608	1.292	0.969	0.614	-	0.973	0.905	0.823	0.670	0.982
I	1.163	0.730	1.006	0.784	-	0.768	0.964	0.901	0.952	0.494
J	1.360	1.300	1.092	1.413	-	1.257	0.927	1.585	1.603	1.623
K	1.059	0.415	1.063	1.700	-	0.992	0.960	1.933	1.798	1.155
L	1.575	0.869	0.984	0.985	-	1.229	0.884	1.008	0.927	1.002
M	1.979	1.065	0.999	0.719	-	1.054	1.039	0.722	0.700	0.585
N	0.636	0.534	0.960	1.779	-	0.822	1.024	2.014	1.856	1.730
O	1.859	0.733	1.035	1.266	-	1.086	1.011	1.189	1.264	0.749
P1	0.758	0.800	1.033	0.852	-	0.853	1.027	1.323	1.407	1.971
P2	0.803	0.931	1.007	0.801	-	0.890	0.959	1.379	1.268	1.826
P3	0.772	0.951	1.001	0.959	-	0.968	1.038	1.221	1.398	1.760
P4	0.837	0.938	1.079	0.867	-	0.987	1.068	1.400	1.422	1.950
P5	0.630	0.876	1.056	0.888	-	0.924	1.011	1.383	1.219	1.800
Ave	1.267	0.789	0.994	1.054	-	0.964	1.088	1.100	1.091	1.088
Cvar(anal)	6.9	2.5	2.7	4.6	-	3.8	2.6	5.0	4.7	2.3
Cvar(rm)	3.466	4.440	0.475	1.997	-	2.150	1.343	0.455	2.202	2.737
Cvar(indiv)	23.146	29.619	0.447	31.065	-	11.261	3.655	35.520	32.960	35.679

	31	32	33	34	35	36	37	38
	TIMP2	tPA	uPA	vWF	MouseIg	RabbitIg	SheepIg	GoatIg
A	1.018	1.077	0.510	2.489	-	-	-	-
B	1.028	1.189	0.713	0.943	-	-	-	-
C	0.776	1.116	0.751	1.583	-	-	-	-
D	1.219	0.686	1.069	1.415	-	-	-	-
E	1.044	0.997	1.067	0.910	-	-	-	-
F	0.992	1.127	0.828	0.874	-	-	-	-
G	0.806	0.872	1.146	0.409	-	-	-	-
H	0.982	0.978	1.007	1.937	-	-	-	-
I	1.369	1.019	1.119	1.263	-	-	-	-
J	0.779	1.388	0.921	0.964	-	-	-	-
K	0.769	0.969	0.829	1.760	-	-	-	-
L	1.176	0.893	1.288	1.002	-	-	-	-
M	0.670	0.944	1.165	0.578	-	-	-	-
N	0.875	0.831	1.457	0.423	-	-	-	-
O	0.685	0.998	0.861	0.659	-	-	-	-
P1	0.969	0.885	0.827	0.723	-	-	-	-
P2	1.133	0.831	0.954	0.877	-	-	-	-
P3	1.007	0.837	0.881	0.723	-	-	-	-
P4	1.068	0.809	0.990	0.762	-	-	-	-
P5	1.034	0.824	0.923	0.610	-	-	-	-
Ave	0.946	0.779	1.307	1.147	-	-	-	-
Cvar(anal)	3.3	2.4	2.9	4.6	-	-	-	-
Cvar(rm)	2.687	1.027	4.036	8.366	-	-	-	-
Cvar(indiv)	15.686	12.983	17.825	38.778	-	-	-	-

Table 2: DMI-derived proteomic data is shown for serum samples prepared from
 venous blood from 15 healthy donors (7 male and 8 female, aged 23 to 37) labelled
 'A' to 'O'. A single serum sample from another individual (male aged 35) was split
 5 into five replicate aliquots (P1 to P5) and also assayed. For each tag, the mean
 normalised fluorescence is shown (to three decimal places). Where no fluorescence
 was detected even in the reference sample alone, a dash is shown. The variance
 components for each tag are broken down and presented: 'Cvar(anal)' is the
 analytical variation from one tag to another within the same experiment. 'Cvar(rm)'
 10 is the repeated measures variation for the 5 replicate aliquots, and is presented net of
 the analytical variation. 'Cvar(indiv)' is the individual-to-individual variation and
 is presented net of both analytical and repeated-measures variation. Proteins with
 higher Cvar(indiv) values contain the most diagnostic information. Dotted boxes
 indicate values outside the calibrated range of the assay (approximately 0.1 to 10
 15 arbitrary units). Black-edged boxes highlight values referred to in the main text.

Furthermore, five of the samples tested were replicate aliquots from the same bleed (P1 to P5, shaded in Table 2). This allows the repeated measures reproducibility (Cvar(rm)) to be assessed. The Cvar(rm) is reported with the analytical variation (Cvar(anal)) subtracted. The median Cvar(rm) for all 31 antibodies for which a signal was detected in the reference sample was 2.7% (range 2.1% to 17.6%) which is slightly inferior to the most robust analytical methods such as NMR for metabonomics (1-2%), but considerably better than any existing proteomic methods, including 2D gel electrophoresis or protein chip microarrays (10-20%).

10 Example 2 : Generation of a large scale DMI antibody library from an unselected phage display library with very high coverage

In example 1, we used a manually constructed small DMI antibody library to illustrate the principle of the approach. However, as with any megaplex technology capable of managing thousands of analytes in parallel, the power of the approach increases with the size of the library. It is not feasible to construct libraries larger than 100 or so components by the manual method, so an alternative is required for large libraries. Furthermore, a manually constructed library will only represent "known" antigens (that is, ones already known or suspected to be present in the test samples). In contrast, a library generated by sub-selection from a phage-display library will be both much larger and likely to contain antibodies to components of the test sample that have never previously been identified.

The prerequisite for successful generation of a large DMI library is a master phage display library with very broad coverage. The higher the number of independent clones composing the master library, the better the resulting DMI library that can be sub-selected from it. The master library can be constructed by any of the methods well known in the art, and examples include the CAT library that contains approximately 10^{13} independent clones, representing at least 10 times the immune diversity of a human subject.

To prepare the large DMI library, an unlabelled aliquot of the reference sample (in our case, the pooled serum from 15 healthy individuals) was coated onto tissue culture plastic (high protein binding plastic) at low protein density (approximately $10\mu\text{g}$ protein per cm^2) to ensure that all, or almost all of the proteins present in the reference sample were bound. A total surface area of about $1,000\text{ cm}^2$ was prepared in this way (with 10mg total protein). The master phage library was then expanded and passed over the plate surface at room temperature for 30 minutes. Unbound phage were washed away thorough with phosphate buffered saline containing 0.1% Tween 20.

The positively selected phage were then released, and the population again expanded. In the second step, the reference sample protein was coated onto tissue culture plastic at very high protein density (10mg of protein per cm^2). With the number of protein binding sites on the plastic severely limiting, many of the rarer proteins will not be represented at all on the plate, while the abundant proteins will be highly represented. The selected phage were then exposed to this surface for 30 minutes at room temperature, and this time the unbound phage were retained and the bound phage were discarded.

This process was repeated a number of times, expanding the phage population, then applying positive selection, expanding the population and performing negative selection and so forth. As the process continued, the redundancy of the library falls, and the bias towards abundant antigens in the reference sample also falls. The bias was monitored as the selection process was iterated: four purified antigens (two abundant (fibrinogen and albumin) and two rare (TGF-beta and PAI-1)) were coated onto ELISA plate wells in 100mM sodium carbonate pH9 at 4°C overnight, then washed and blocked using 5% sucrose/5% Tween in phosphate buffered saline. After washing the wells again (in phosphate buffered saline + 0.1% Tween) a serial dilution of the selected library was applied to each antigen. This was allowed to bind for 30 minutes at room temperature, then the wells were washed, and the bound phage detected with an anti-phage coat protein antibody labelled with horseradish peroxidase. After further washes, the amount of bound enzyme was quantitated

using the substrate K-BLUE. The dilution of the library that yielded half maximal signal on each antigen was then determined (with undiluted library assigned the arbitrary concentration of 1 unit). The bias of the library was calculated as the mean for the two abundant antigens divided by the mean for the two rare antigens. The bias of the subselected DMI library as we performed four iterations of positive and negative selection are shown in Figure 4.

This example demonstrates that it is possible to generate a large DMI library with low redundancy and low bias which could be limiting dilution cloned in microtitre plates to generate a tagged library similar to the one used in example 1 but with 10,000 to 100,000 individual components.

Example 3: Immunomics using a small-scale carbohydrate antigen library

As the first step, an antigen library must be assembled. For this pilot-scale experiment, the library was manually constructed by dispensing individually synthesised and purified carbohydrate antigens into wells of a 96 well plate. Twenty four different oligosaccharide sequences were commercially available (Glycorex) coupled to serum albumin (Table 3). Serum albumins (bovine or human origin) without any carbohydrate attached were used as control library components dispensed into 2 further wells. In each well, approximately 100µg of protein/oligosaccharide conjugate was dispensed.

Table 3

Tag	Antigen	Conjugate	Carrier	CVar
1	Glcβ-O-spacer	B-1001	BSA	2.1
2	Galβ-O-spacer	B-1002	BSA	2.3
3	Manα-O-spacer	B-1003	BSA	1.9 (M)
4	Galβ1-4Glcβ-O-spacer	B-1004	BSA	4.8
5	Galβ1-4GlcNAcβ-O-spacer	B-1005	BSA	3.0
6	Glcα1-6Glcα1-4Glcβ1-4Glcβ-O-spacer	B-1007	BSA	-
7	Galα1-4Galβ1-4Glcβ-O-spacer	B-1017	BSA	2.2
8	Galα1-4Galβ1-4GlcNAcβ-O-spacer	B-1010	BSA	2.6
9	Galα1-4Galβ-O-spacer	B-1011	BSA	2.1

10	Gal β 1-3GlcNAc β -O-spacer	B-1012	BSA	2.4
11	Di-Man α 1-6(α 1-3)Man α -O-spacer	B-1014	BSA	-
12	GalNAc β 1-3Gal α -O-spacer	B-1015	BSA	2.7
13	GalNAc β 1-4Gal β -O-spacer	B-1016	BSA	2.2
14	GalNAc β -O-spacer	B-1018	BSA	2.1
15	GalNAc α 1-3(Fuc α 1-2)Gal β -O-spacer	B-1019	BSA	6.1
16	Gal α 1-3(Fuc α 1-2)Gal β -O-spacer	B-1020	BSA	4.4
17	Gal α 1-3Gal-O-spacer	B-1008	BSA	2.4
18	Gal α 1-3Gal β 1-4GlcNAc β -O-spacer	B-1009	BSA	2.5
19	Gal α -O-spacer	H-1021	HSA	3.3
20	Gal α 1-2Gal-O-spacer	H-1022	HSA	3.2
21	Gal α 1-3Gal β 1-4GlcNAc β 1-3Gal β 1-4Glc-O-spacer	H-1025	HSA	2.3
22	Gal α 1-4Gal-O-spacer	H-1026	HSA	2.8
23	Gal α 1-3GalNAc α -O-spacer	H-1030	HSA	3.7
24	Gal β 1-3GalNAc α -O-spacer	H-1031	HSA	3.2
25	None	Glycorex	BSA	6.9 (M)
26	None	Glycorex	HSA	-

Table 3: The glycoconjugate antigens that were selected to generate the small manual DMI library for immunomics are shown above. ‘Tag’ numbers represent the position of the library component in the output vector (and is not the code of the tag, which is more complex). ‘Antigen’ represents the carbohydrate sequence in the conjugate. ‘Conjugate’ represents the source of the particular conjugate used – all the catalog codes refer to the Glycorex catalog. ‘Carrier’ indicates the carrier protein to which the carbohydrate antigens are conjugated, where BSA represents bovine serum albumin and HSA represents human serum albumin. Unconjugated aliquots of the same batch of these proteins were used as controls on tags 25 and 26. ‘Cvar’ is the coefficient of variation for reading multiple tags of the same code in the same experiment. The Cvar is the mean of the Cvar for the pan-IgG (FITC) vector and the IgM (rPE) vector, except where stated when too little IgG bound to the antigen to be quantified. A dash indicates that neither Ig class bound to the antigen to any significant degree. Note that the Cvar reported is the mean from 15 different individuals, to reflect the varying signal bound to each tag which results in a varying analytical Cvar from individual to individual (in contrast to Table 1, where the analytical Cvar depends on the average signal from all of the individuals, represented by the reference sample).

The antigen library was then tagged, using aluminium bar code tags, exactly as described in example 1 for an antibody library. Since the oligosaccharide antigens were carried on protein scaffolds, the same chemistry that is used to bind antibody protein to the aluminium, also achieves attachment of the oligosaccharide/protein
5 conjugates. A different pool of aluminium bar coded tags was dispensed into each well (about 10^4 individual tags in each pool). At the end of the tagging reaction, the tags were harvested and washed in phosphate-buffered saline by gentle ultrafiltration, and resuspended in 100 μ l per well of phosphate-buffered saline. All the wells were
10 then combined to yield approximately 2ml of library containing a total of 2×10^5 individual tags at 100,000 tags per ml.

In the second step, serum samples from 15 healthy volunteers were dispensed at 20 μ l per sample directly into V-bottom microtitre plate wells. 20 μ l of the library was then
15 added (approximately 2,000 individual tags, representing a 100-fold excess over the number of individual components of the library). Non-ionic detergent (Tween 20 at 0.1% vol/vol final concentration) was also added to the reaction mixture to improve the specificity of antibody binding, and lower the background. The plate was then sealed and the reaction mixed thoroughly, and incubated at room temperature with
20 continual agitation for 15 minutes.

At the end of the incubation, the tags were harvested and washed by gentle ultrafiltration over a vacuum manifold, and phosphate-buffered saline containing 0.1% Tween 20 was used throughout as the wash solution. The beads were then
25 resuspended in 50 μ l of phosphate-buffered saline with 0.1% Tween 20 and each of the WHO standard mouse monoclonal anti-human Ig class specific antibodies labelled with a different fluorochrome. For this experiment, we used the anti-pan IgG antibody labelled with FITC and the anti-IgM antibody labelled with TRITC. Each of the detection antibodies was present at 5 μ g/ml final concentration. The plate
30 was then sealed and mixed, before being incubated at room temperature with continual agitation for 15 minutes.

As the third step, for detection of the antibodies a fluorescence microscope was used. The reaction from each well in turn was dispensed onto a standard glass microscope slide in a well about 1cm in diameter inscribed using a PAP pen. A coverslip was placed over the slide and sealed to prevent evaporation using clear nail varnish. The slide was then placed under a fluorescence microscope, and the bar coded tags located, one at a time, under direct illumination. As each tag was located, its binary code was read and logged. The amount of fluorescence in the fluorescein channel and rhodamine channel were then determined using an automated filterwheel changer. The two separate fluorescence readings were then recorded together with the bar code for each tag. Where more than one tag was located in each reaction with the same binary code, the fluorescence readings from the two (or more) identical tags were averaged prior to reporting the immunomic profile vector. Approximately 500 individual tags were read for each reaction. Using a manual microscope system, this take approximately one hour per sample analysed. However, automated systems do exist for reading the fluorescence bound to each bar coded tag under a microscope. Alternatively, the tags could be read using an appropriate flow cytometer (see example 1).

The resulting vectors for the 15 individuals are shown in Table 4. For each antigen tag, there are two columns: the left-hand column contains the pan-IgG parameter and the right-hand column contains the IgM parameter. These vectors represent the IgG/M immunomic profile (focussed on carbohydrate antigens) for each of the individuals tested, and can be used for various investigational or analytical purposes.

25 Table 4

	1		2		3		4		5		6		7	
									Nac-lacA		GlyStor		Pk	
A	3	141	2	0	0	0	0	10	35	23	0	0	140	103
B	21	116	1	1	0	13	1	6	24	57	0	0	2	7
C	14	30	6	39	0	0	4	13	40	108	0	0	107	410
D	13	45	2	42	0	0	0	2	36	7	0	0	119	125
E	11	20	6	33	0	0	3	7	48	43	0	0	0	68
F	1	113	3	14	0	3	282	44	35	151	0	0	1	31
G	22	52	4	552	1	2	25	15	53	52	33	244	75	134
H	7	30	8	2	0	0	4	15	55	70	0	0	142	99
I	23	43	3	1	0	1	0	10	73	189	0	0	53	86
J	2	94	2	10	3	1	1	27	35	68	0	0	238	113
K	21	32	1	11	0	0	96	27	101	200	0	1	62	321
L	5	48	2	15	0	0	94	54	20	84	0	0	201	231
M	5	39	2	12	0	0	0	11	97	43	0	0	137	371
N	11	34	4	6	0	0	68	43	28	42	0	0	142	122
O	6	37	4	6	0	0	1	31	28	46	0	0	221	960
P1	3	33	6	13	0	1	4	2	37	68	0	0	53	2
P2	3	42	5	15	0	0	3	1	42	60	0	0	68	2
P3	4	47	5	19	0	2	3	2	44	68	0	0	17	2
P4	3	37	5	15	0	1	3	2	42	61	0	1	69	2
P5	4	39	5	16	0	1	4	2	38	67	0	0	60	2
Median	11	43	3	11	0	0	3	15	36	57	0	0	119	122
Cvar(anal)	2.2	2.1	2.1	2.5	-	11.9	5.5	4.1	3.3	2.7	-	-	2.2	2.2
Cvar(rm)	13.9	11.2	6.5	11.5	-	49.5	10.6	9.0	4.0	3.4	-	-	37.8	8.2
Cvar(indiv)	54	52	52	267	-	185	180	63	46	68	-	-	30	103

	8		9		10		11		12		13		14	
	P1		EColiR											
A	29	32	87	454	3	4	0	0	6	8	5	9	1	4
B	136	242	6	59	3	8	0	1	5	10	2	19	13	4
C	62	87	41	0	1	6	0	3	8	153	6	5	21	32
D	94	109	15	5	6	3	0	0	5	33	7	9	1	2
E	211	581	5	15	2	20	0	0	4	6	4	22	2	2
F	176	146	46	5	1	2	0	0	6	9	3	14	0	3
G	74	102	2	3	7	3	0	0	4	29	5	17	1	4
H	33	78	65	41	2	4	0	0	4	23	4	7	3	2
I	71	32	7	363	4	6	0	0	4	16	5	8	15	20
J	41	293	45	361	2	3	0	0	6	12	5	13	3	4
K	27	32	4	4	8	36	0	0	14	12	4	13	1	2
L	63	93	13	150	1	6	0	0	8	9	4	8	1	6
M	91	57	96	18	11	7	0	0	10	13	2	9	3	10
N	60	178	12	1	9	4	0	0	4	51	3	5	4	20
O	100	68	0	1	1	21	0	0	2	9	0	2	3	1
P1	103	143	56	21	3	6	0	0	4	38	1	10	4	13
P2	97	157	52	16	3	5	0	0	3	32	2	10	3	17
P3	104	155	48	18	1	5	0	0	4	40	1	12	5	14
P4	109	155	47	21	3	7	0	0	4	31	1	13	5	11
P5	102	160	46	18	2	3	0	0	3	33	1	11	3	16
Median	71	93	13	15	3	6	0	0	5	12	4	9	3	4
Cvar(anal)	2.2	3.0	2.0	2.3	2.7	2.1	-	-	3.0	2.5	2.1	2.2	2.1	2.1
Cvar(rm)	2.0	1.2	6.3	9.2	34.6	26.4	-	-	12.2	8.9	35.2	9.9	22.9	14.7
Cvar(indiv)	59	97	100	149	44	78	-	-	35	130	7	40	106	101

	15 A		16 B		17 Di-aGal		18 Tri-aGal		19		20		21 Pentagal	
A	252	557	293	296	81	133	108	92	3	26	6	59	77	68
B	198	1098	461	607	119	62	830	456	4	10	14	21	465	696
C	1	127	569	113	67	31	46	30	1	22	84	32	43	881
D	438	231	213	458	33	29	138	44	2	37	25	13	18	324
E	0	15	147	1436	47	39	1160	124	4	467	146	148	436	245
F	0	38	336	209	82	108	32	161	5	34	5	54	58	89
G	69	1664	0	3	16	19	40	67	6	40	26	12	34	58
H	7	11	289	469	46	72	242	287	2	20	3	34	82	39
I	552	208	119	991	13	84	161	132	5	11	27	99	65	218
J	1	4	460	526	35	127	149	536	4	30	3	12	12	94
K	0	46	238	672	12	27	67	87	6	16	30	38	29	475
L	297	794	301	219	104	75	553	148	5	44	2	102	25	264
M	0	43	262	816	10	127	69	1317	5	27	6	54	24	405
N	0	3	290	655	64	40	81	562	3	12	1	44	45	78
O	360	288	452	200	422	135	409	589	7	17	335	422	5	482
P1	278	462	221	627	64	117	162	442	13	20	5	49	70	268
P2	256	398	292	556	82	109	178	409	11	23	7	42	73	242
P3	292	450	165	691	73	102	155	471	11	27	6	27	66	209
P4	291	426	244	603	79	116	159	477	12	26	6	46	71	253
P5	258	511	268	617	89	92	177	504	10	26	5	48	84	257
Median	7	127	290	469	47	72	138	148	4	26	14	44	43	245
Cvar(anal)	5.5	6.7	4.2	4.6	2.3	2.4	2.4	2.6	3.6	3.0	3.2	3.3	2.3	2.4
Cvar(rm)	0.8	2.7	16.2	3.3	9.9	7.3	4.0	5.3	6.4	8.8	11.2	18.0	7.0	6.8
Cvar(indiv)	125	134	30	66	120	48	116	103	31	200	172	114	146	77

	22		23		24		25 BSA		26 HSA	
A	37	311	4	17	19	177	0	0	0	0
B	14	135	9	39	32	31	0	3	0	0
C	13	1915	51	194	51	31	0	17	0	0
D	4	7	37	50	7	16	0	2	0	0
E	107	608	68	552	92	166	0	1	0	0
F	20	6	13	318	14	20	0	9	0	0
G	74	12	16	47	97	14	0	2	0	0
H	22	15	5	104	39	8	0	3	0	0
I	40	4	147	38	33	144	46	191	0	0
J	34	299	22	113	107	307	0	1	0	0
K	11	10	18	53	39	59	0	0	0	0
L	19	1	12	65	29	39	0	3	0	0
M	4	4	11	76	53	35	0	1	0	0
N	29	2	109	262	126	34	0	1	0	0
O	22	54	154	175	84	126	0	0	0	0
P1	4	14	38	209	26	172	0	3	0	0
P2	3	11	46	248	22	174	0	2	0	0
P3	4	10	52	238	19	188	0	2	0	0
P4	3	13	59	258	25	169	0	3	0	0
P5	4	12	54	250	23	170	0	4	0	0
Median	22	12	18	76	39	35	0	2	0	0
Cvar(anal)	2.7	2.9	2.8	4.6	3.4	3.0	-	6.9	-	-
Cvar(rm)	12.5	10.3	13.4	3.3	8.5	1.4	-	23.0	-	-
Cvar(indiv)	77	208	98	95	56	102	-	282	-	-

Table 4: DMI-derived immunomic data is shown for serum samples prepared from venous blood from 15 healthy donors (7 male and 8 female, aged 23 to 37) labelled 'A' to 'O'. A single serum sample from another individual (male aged 35) was split into five replicate aliquots (P1 to P5) and also assayed. For each tag, the mean fluorescence bound is shown for pan-IgG (FITC) in the left-hand column and IgM (rPE) in the right-hand column. The variance components for each tag are broken down and presented: 'Cvar(anal)' is the analytical variation from one tag to another within the same experiment. 'Cvar(rm)' is the repeated measures variation for the 5 replicate aliquots, and is presented net of the analytical variation. 'Cvar(individ)' is the individual-to-individual variation and is presented net of both analytical and repeated-measures variation. Proteins with higher Cvar(individ) values contain the most diagnostic information. Note that many of the tags yielded an approximately log-normal distribution, and that it would be appropriate log-transform the data prior to calculation of more accurate variance components. Furthermore, the data is heavily influenced by outliers – the impact of these outliers would be reduced by transformation, but Winzorising may be more appropriate once larger immunomic datasets were collected.

In this example, we noted that about half the individuals had high levels of IgG (and also IgM) antibodies bound to tag 15 (values boxed in Table 4). This tag has the carbohydrate structure representing the A blood group antigen bound to it. The individuals with low levels of antibody must themselves express the A antigen and are either A or AB blood group. The individuals with high levels of antibody must not express the A antigen and are either O or B blood group. In fact, the same reasoning can be applied to the data from tag 16 which has the carbohydrate structure representing the B blood group antigen bound to it. From these two columns it is possible to determine that individual F is blood group A, while individual G is blood group B and individual L is blood group O. The same deductive process can be applied to all the individuals studied.

As for the use of DMI in proteomics (example 1), we have performed a full analysis of the sources of variation within the immunomic dataset (Tables 3 & 4). Firstly, we

have assessed the analytical reproducibility of the method ($Cvar(anal)$) calculated from the range of fluorescence readings from different tags with the same code in the same experiment. Unlike the proteomic analysis the $Cvar(anal)$ varies from individual to individual because the absolute level of signal varies from individual to individual. The $Cvar(anal)$ values reported are therefore the mean value for the 15 individuals studied. The analytical reproducibility is excellent (below 5% for most tags, superior to individual immunoassays).

Furthermore, five of the samples tested were replicate aliquots from the same bleed (P1 to P5, shaded in Table 4). This allows the repeated measures reproducibility ($Cvar(rm)$) to be assessed. The $Cvar(rm)$ is reported with the analytical variation ($Cvar(anal)$) subtracted. The median $Cvar(rm)$ for all 22 antigen tags for which a signal was detected in more than one test sample was 9% (range 0.8% to 49.5%) which is somewhat inferior to the application of DMI to proteomics. However, the reason for this lies in part in the very low signals which were obtained for many individuals on many of the tags – low signal, near the detection limit of the technique, is always detected with lower repeated measures reproducibility. However, the $Cvar(individ)$, which represents the true individual-to-individual variance component is larger for the immunomic vectors than for the proteomic vectors (compare Table 4 with Table 2). This is the variance component which is useful for diagnostic modelling. Consequently, the true diagnostic utility of the test, which is approximated by $Cvar(rm)/Cvar(individ)$ is very similar in the two applications of DMI.

It is important to note that the signal for each of the tags approximates a log-normal distribution, and that there are also a number of extreme outliers in the dataset. Consequently, a more thorough analysis would require log transformation (and possibly Winsorising) of the dataset prior to further investigation of the X-matrix.

Example 4: Preparation of a large peptide antigen library for DMI-based immunomics

To generate a large scale peptide antigen library, the following strategy was adopted: nine amino acid peptides were chosen to represent the master library. However, there are 20^9 (about 5×10^{11}) sequence variants that compose this master library - many times too many for them all to be uniquely represented in the DMI antigen library. Therefore, to generate a library of manageable proportions, the amino acids were grouped into 4 groups of 5 based on similarity of properties (dominantly, charge and hydrophobicity). The groups selected were: GROUP 1 (charged) Arg, Lys, His, Asp, Glu; GROUP 2 (small hydrophobic) Gly, Ala, Leu, Ile, Val; GROUP 3 (large hydrophobic) Met, Phe, Pro, Tyr, Trp and GROUP 4 (hydrophilic) Ser, Thr, Asn, Gln, Cys. Alternative groupings could also be adopted, and would yield subtly different libraries that would still be suitable for immunomics. An equimolar mixture of the five amino acids within the group was then treated as a single reagent for combinatorial solid phase synthesis. There are, therefore, now just 4^9 possible components to the library (262,144 components). Note, however, that each "component" is not a single peptide sequence but a mixture of 5^9 (1.6 million) possible sequence variants - however, because of the grouping of the amino acids, related sequences are likely to fall within the same component pool.

The 262,144 component pools were synthesised by solid-phase synthesis using methods well known in the art. Briefly, each group of amino acids were coupled onto batches of solid phase resin. Each batch of coupled resin was then divided into four, and reacted with one of the four groups of amino acids, using appropriately protected amino acids. This process was then repeated, until a total of 262,144 batches of resin had been generated. Each was then cleaved and deprotected in parallel to yield 690 microtitre plates (384 wells per plate) each containing approximately 1mg of peptide.

To each individual well, a different aluminium bar code tag pool was added approximately 10^6 identical individual tags in each case), and the peptide was allowed to bind to the tags. The tags were then removed and washed by gentle ultrafiltration, and resuspended in 100µl of phosphate-buffered saline. All the components of the library were then combined, to yield 26 litres of pooled library containing

approximately 10^{12} individual tags (approximately 10^7 tags per ml). This library was then concentrated by gentle ultrafiltration to a final volume of 250ml (10^8 tags/ml) which was then suitable for use at 20 μ l per sample as in example 3 (allowing a total of more than 12,500 samples to be measured with this library).

.5

This example demonstrates that it is possible to generate a very large antigen library capable of generating a high data density immunomic vector that contains information about antibodies recognising all possible 9 amino acid peptide antigens (every antigen is present, even though not every one is individually distinguishable as a separate library component). This library can be used to obtain an immunomic profile vector containing 2,359,296 individual datapoints for each individual in a procedure taking 30 minutes, exactly as described for the small carbohydrate antigen library in example 3.

10

CLAIMS

1. A method of determining the relative abundance of a plurality of proteins in a test sample compared to a reference sample, the method comprising:
 - 5 (a) providing a reference sample comprising a plurality of labelled proteins;
 - (b) incubating a plurality of tagged antibodies capable of binding components of the reference sample with (i) a mixture of the labelled reference sample and the test sample and (ii) the reference sample alone, under conditions
10 suitable for the binding of said antibodies to their targets;
 - (c) comparing the amount of labelled protein bound to individual antibody tags in the presence and absence of the test sample.
2. A method according to claim 1 wherein said test sample and reference
15 sample are mixed in equal volumes.
3. A method according to claim 1 or 2 wherein said antibodies are tagged with aluminium bar codes or dye impregnated beads
- 20 4. A method according to any one of the preceding claims wherein each tag is linked to a single antibody species.
5. A method according to any one of claims 1 to 3 wherein each tag is linked to more than one species of antibody.
25
6. A method according to claim 5 wherein each of said antibody species linked to a tag binds the same protein.
7. A method according to any one of claims 1 to 5 wherein each of said
30 plurality of tagged antibodies binds a different protein.

8. A method according to any one of the preceding claims wherein from 10^{11} to 10^{15} antibody molecules are bound to each tag.

5 7. A method according to any one of the preceding claims wherein said reference sample is obtained from the same tissue and/or organism as said test sample.

8. A method according to any one of the preceding claims wherein said reference sample is formed by pooling a plurality of test samples.

10

9. A method according to any one of the preceding claims wherein said proteins in the reference sample are labelled with one or more fluorescent dyes.

10. A method according to any one of the preceding claims wherein said binding is quantified by flow cytometry.

15

11. A method of detecting a plurality of immunoglobulins in a test sample, the method comprising:

- (a) providing a plurality of tagged antigens;
- 20 (b) incubating said tagged antigens of (a) with said test sample, under conditions suitable for the binding of any immunoglobulins present in said test sample to their targets;
- (c) incubating said mixture of (b) with one or more labelled antibodies capable of binding specifically to immunoglobulins;
- 25 (d) measuring the amount of labelled antibody bound to each tagged antigen.

25

12. A method according to claim 11 wherein said plurality of antigens comprises oligopeptides and/or oligosaccharides.

30

13. A method according to claim 11 or 12 wherein each of said antigens comprises a different tag.

14. A method according to any one of claims 11 to 13 wherein said labelled antibodies comprise antibodies specific to two or more immunoglobulin subclasses.

5 15. A method according to claim 14 wherein said antibodies specific to each immunoglobulin subclass comprise a different label.

16. A method according to claim 14 or 15 wherein said immunoglobulin subclasses are selected from IgG1, IgG2, IgG3, IgA, IgD, IgE and IgM.

10 17. A method according to any one of claims 11 to 16 further comprising the step of quantifying the amount of each immunoglobulin subclass that binds each tagged antigen.

15 18. A method according to any one of claims 11 to 17 wherein the amount of labelled antibody bound to each tagged antigen is measured by flow cytometry.

19. A method of reducing the redundancy and bias of an antibody-expressing phage library comprising:

20 (a) providing two surfaces to which a sample of antigens is bound wherein said antigens are bound to the second surface at a higher density than to the first surface;

(b) exposing a phage display library to a first surface of (a) under conditions suitable for antibody binding and selecting phage bound to said surface;

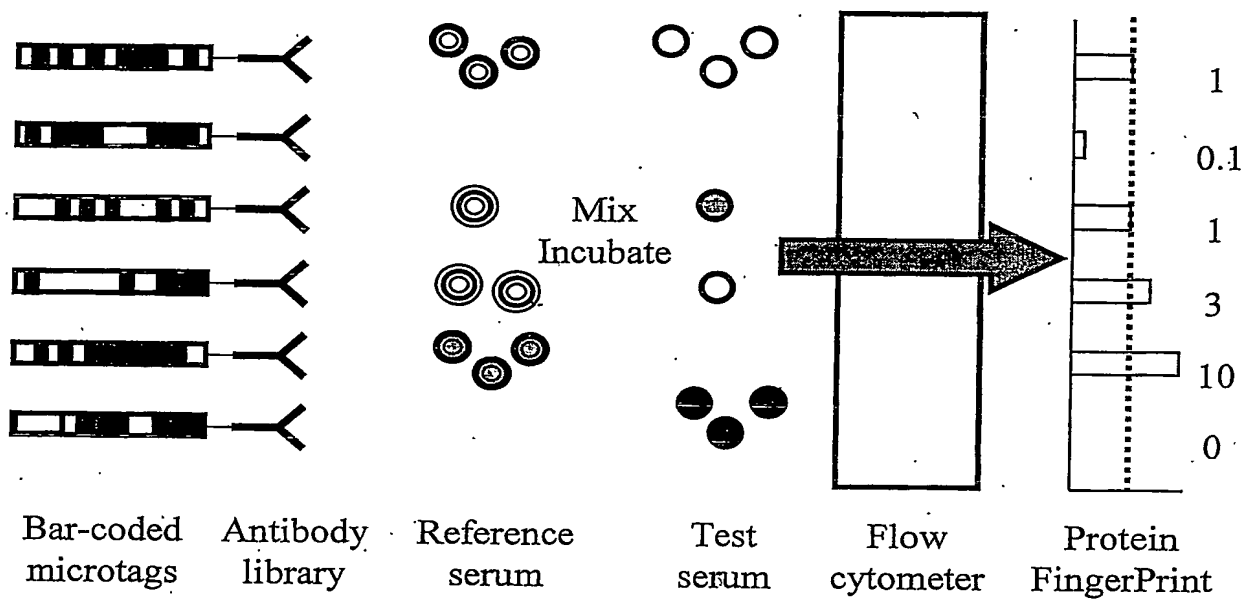
25 (c) exposing said selected phage of (b) to a second surface of (a) under conditions suitable for antibody binding and selecting phage not bound to said surface;

(d) optionally further selecting said phage of (c) according to steps (b) and (c) one or more times;

30 thereby obtaining a library of antibody-expressing phage which has reduced redundancy and/or bias characteristics compared with the original library.

20. A method according to claim 1 wherein said plurality of antibodies is an antibody-expressing phage library produced according to the method of claim 19.

A



B

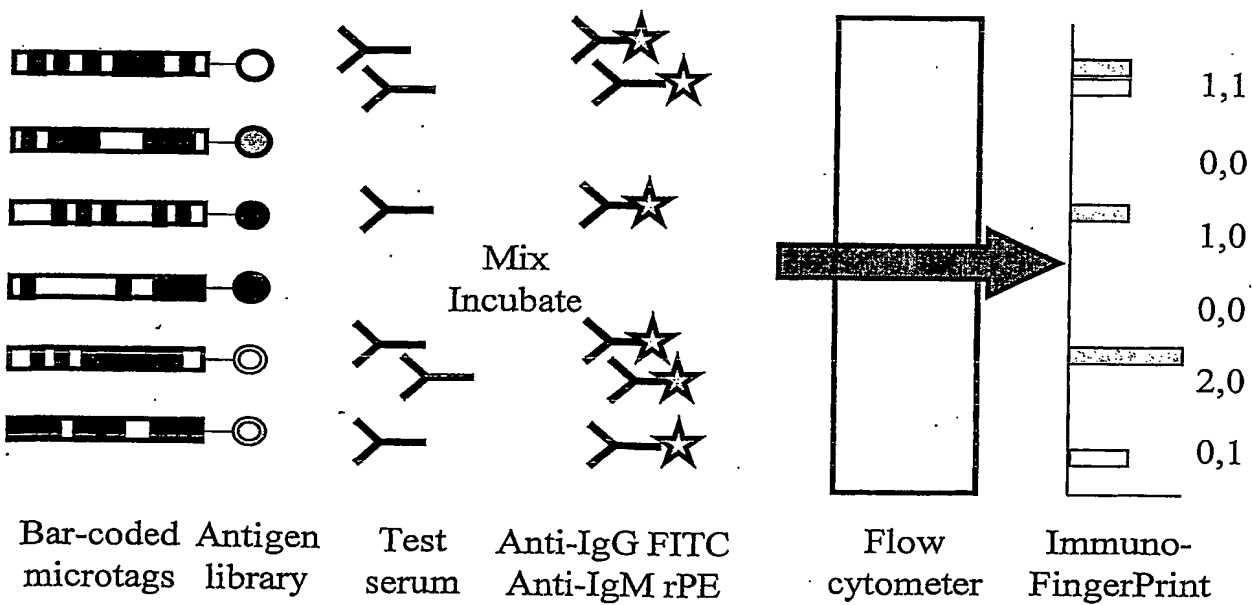


Figure 1

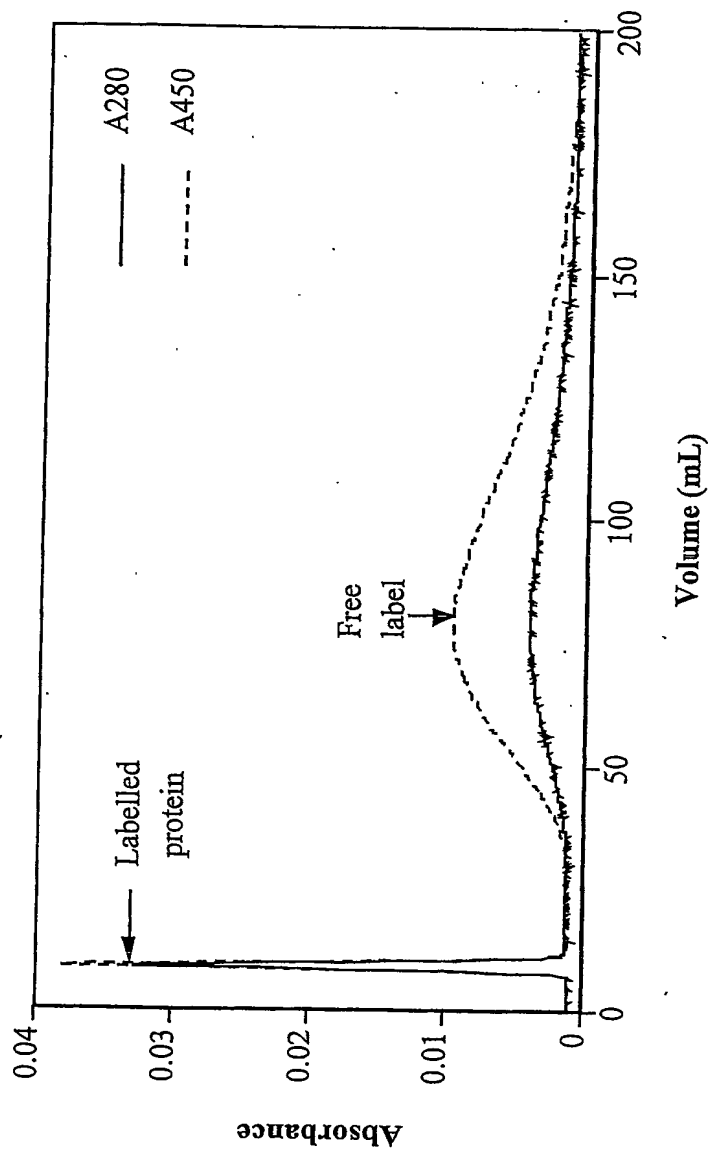


Figure 2

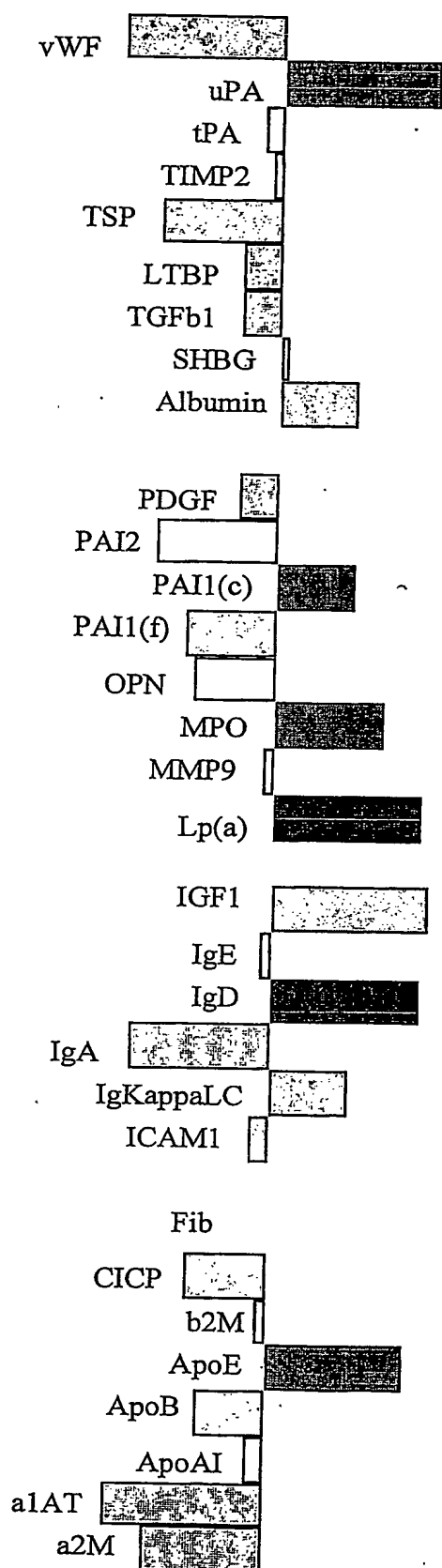


Figure 3

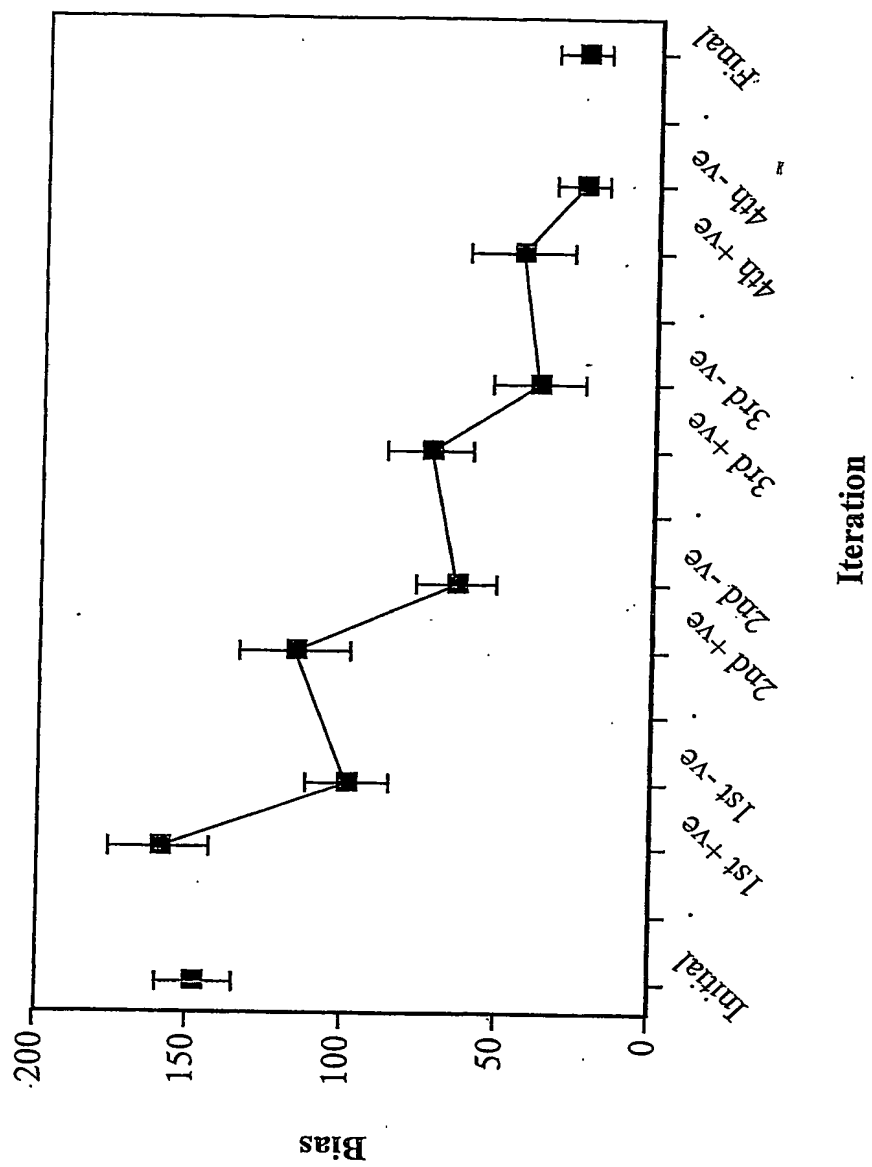


Figure 4

**This Page is Inserted by IFW Indexing and Scanning
Operations and is not part of the Official Record**

BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images include but are not limited to the items checked:

- ☐ **BLACK BORDERS**
- ☐ **IMAGE CUT OFF AT TOP, BOTTOM OR SIDES**
- ☐ **FADED TEXT OR DRAWING**
- ☐ **BLURRED OR ILLEGIBLE TEXT OR DRAWING**
- ☐ **SKEWED/SLANTED IMAGES**
- ☐ **COLOR OR BLACK AND WHITE PHOTOGRAPHS**
- ☐ **GRAY SCALE DOCUMENTS**
- ☐ **LINES OR MARKS ON ORIGINAL DOCUMENT**
- ☐ **REFERENCE(S) OR EXHIBIT(S) SUBMITTED ARE POOR QUALITY**
- ☐ **OTHER: _____**

IMAGES ARE BEST AVAILABLE COPY.

As rescanning these documents will not correct the image problems checked, please do not report these problems to the IFW Image Problem Mailbox.